Econophysics and Agent-Based Modeling of Financial Market

FENG LING

NATIONAL UNIVERSITY OF SINGAPORE

2013

Econophysics and Agent-Based Modeling of Financial Market

FENG LING

(B.Sc.(Hons), National University of Singapore)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

NUS GRADUATE SCHOOL FOR INTEGRATIVE SCIENCES AND ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2013

Declaration

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Feng Ling

17 June 2013

Acknowledgements

Many people have contributed to this dissertation in different ways, and it is my esteemed pleasure to thank them, without whom the thesis could never been possible.

First and foremost, I would like to express my greatest gratitude to my Ph.D supervisor professor Li Baowen for his invaluable supervision, guidance and patience. It has been a privilege to be his Ph.D student, and I am extremely thankful for his inspiring discussions, and sharing of knowledge. At the same time, I have to thank the academic freedom and independence he gave me to think and explore an exciting territory of highly interdisciplinary research area between physics and economics. His enthusiasm and insight have motivated me throughout the whole course my Ph.D studies.

I would also like to thank my Thesis Advisory Committee (TAC) members professor Gong Jiangbin, professor Joseph Cherian and professor Chen Ying, for their valuable advices and critical comments on my research project. Since they come from different departments, it has not been easy for them to communicate and coordinate as my advisory committee. But they have spent their valuable time and effort in making this possible.

I am extremely grateful to Professor H. Eugene Stanley from Boston University, for being a very helpful collaborator and caring mentor. The time I have spent in his group at Boston gave me a great deal of exposure to the frontiers in this research area. His insight and guidance are the most rewarding experience to me during that period, not to mention his hospitality and warm-heartedness as a personal friend who made that part of my life a rather unforgettable experience.

My thesis would also be in no part possible without the help of my collaborator professor Boris Podobnik, who has also been a very pleasant friend and mentor, providing me with insightful perspective on the work as well as a broader picture on how to carry out high quality research. Professor Tobias Preis has also been a valuable collaborator and friend who helped me along the way, with his expert opinions and techniques.

Special thanks go to my group members and friends Liu Sha, Zhang Xun, Zhu Guimei, Yang Lina, Ren Jie, Ma Jing and Qiao Zhi for sharing the joys and pains with me. Their company during my candidature played an irreplaceable part of life. I also thank professor Yang Huijie for his insightful discussions during the early stage of my candidature.

Finally, my upmost appreciation goes to my beloved parents, who provide their unconditional love and support in every aspects of my life.

Table of Contents

Summary

Econophysics is an interdisciplinary research area where physicists apply their thinkings and methods to economics. The field is motivated by extensive empirical observations made available through the growing volume of economic data. This study focuses on one particular aspect of econophysics - agent-based modeling of financial market. There are a number of universal patterns found in financial time series called 'stylized facts'; among them are fat-tail distributions and long memory in volatilities. These patterns cannot be explained by the theory of rational agents or efficient market hypothesis. Additionally, the market agents interact with each other to give rich phenomena including social learning and herding. Hence in the past decades, there has been a growing trend of simulating the financial market based on the interactions of agents with different behavioral rules, and such models are called agent-based models.

Empirical evidence indicates that technical traders dominate trading activity with shorter holding period compared to fundamental investors. Hence daily price fluctuations are influenced much more by technical traders. Empirical and theoretical evidence suggests that agents tend to have similar opinions after large price fluctuations, and diverged opinions after tranquil market conditions. Such a mechanism

can be incorporated into an agent-based model to simulation price dynamics. With a realistic set of parameters, the model produces power-law tail distribution in returns with numerical accuracy, and the result is found to be robust with different parameter values. In addition, the model is able to capture the power-law tail distribution in daily number of trades at the same time, signifying a sound mechanism underlying the model. By incorporating the intraday seasonality in trading volumes, extension of the model to intraday returns produces several empirical statistical patterns at high frequencies.

Theoretical analysis on the agent-based model is carried out in line with Autoregressive Conditional Heteroskedasticity (ARCH) model formulation. With the additional empirical evidence showing heterogeneous investment horizons of agents, an extended stochastic volatility model is derived and it is able to produce long memory of volatility found in financial time series, with quantitative accuracy. Mathematical analysis leads to a scaling relation between the heterogeneity of investment horizons and decay speed of volatility. The analysis provides a behavioral interpretation of the long-term memory of absolute and squared price returns: they are directly linked to the way investors evaluate their investments by applying technical strategies at different investment horizons, and this quantitative relationship is found to be in agreement with empirical findings.

By drawing an analogy to phase transition in physics, one-to-one correspondence between financial market and statistical mechanics is made, and the long-range correlation in time scale directly gives rise to the apparent critical phenomenon in financial market. Comparing the stochastic volatility model with other ARCH family models, this work gives an explicit interpretation to the general formulation of ARCH models in terms of agent behaviors, and provides a new avenue for calibrating parameters based on empirical behaviors rather than statistical fitting.

Publications

1. Ling Feng, Baowen Li, Boris Podobnik, Tobias Preis, and H. Eugene Stanley, Linking agent-based models and stochastic models of financial markets. Proceedings of the National Academy of Sciences, 109(22):83888393, 2012.

2. Zeyu Zheng, Boris Podobnik, Ling Feng, Baowen Li, Changes in Cross-Correlations as an Indicator for Systemic Risk, Scientific Reports 2, Article number: 888, 2012.

List of Tables

List of Figures

Chapter 1

Introduction

The term *econophysics* refers a relatively new discipline that aims to deal with economic systems using the thinkings and approachs from physics. While it is still arguable what exactly the definition of econophysics is, the way physicists deal with such a complex system of economics has definitely provided some new perspectives. While economic theories usually assume rational market participants (agents) with certain utility functions, new evidence points out the fact of abundant irrational behaviors among them. Additionally, the agents interact with each other to give rich phenomena including social learning and herding, as opposed to the idea of 'representative agents' who only look at price and utility. The complexity of such interactions has not been paid much attention until recently. Physicists have a long history of dealing with large number of interacting particles in the field of statistical mechanics. Although it is a little far-stretched to say human beings can be treated as mechanical particles in physical systems, it does not negate some meaningful contribution from physicists in the field of economics, and such endeavor is called agent-based modeling.

This study only focuses on the study of financial market, which is possibly the most studied among all areas of econophysics, due to the availability of large amount of financial data at high frequencies in the recent decades and the market's essential role in modern economy. This field has been motivated by various empirical facts found empirically in financial time series throughout the past decades, first of which are fat-tail and volatility clustering by B. Mandelbrot in 1963 [1]. In the past two decades, due to the availability of large amount of financial data, more statistical facts were discovered in greater detail. H. E. Stanley has found power-law tails in the distribution of returns [2], trading volumes and number of trades [3]. Such power-law tails follow largely invariant tail exponents across different markets in the world. In statistical mechanics, similar universality phenomenon exists in phase transition, and through renormalization theory such universality has been successfully explained in terms of the interaction among particles regardless of the size of interaction strength. This led physicists to wonder if there is a dominant interaction mechanism underlying market dynamics that give rise to those universal power-law exponents. Similar to particles, agents do interact with each other in market with certain behaviors traits. If we model market based on interacting agents, the models are named agent-based models.

In this work, agent behaviors are gathered through empirical evidence and theoretical intuition, and an agent-based model of financial market is constructed. The mathematical analysis of the model is carried out using ARCH [4] formulation, and extension is made to give a stochastic volatility model empirically calibrated from market data. This thesis is presented in seven chapters. In this chapter, an overview is presented on the existing empirical and theoretical studies. Chapter 2 describes the data and methods used in this study, and Chapter 3 presents the result of some empirical data analysis. Chapter 4 describes a new agent-based model (ABM) of financial market, with some investigation on empirical behaviors based on market data. Chapter 5 extends the ABM to a stochastic time series model with additional empirical behaviors, and a detailed study into this new stochastic volatility model is presented. Chapter 6 draws relation between our stochastic volatility model and existing models, and presents a variation of our model. Chapter 7 concludes the whole study and proposes further work.

1.1 Stylized facts in financial time series

Among all the economic time series, financial time series is possibly the most comprehensively documented. This allows academic community to carry out thorough investigations on financial markets. Many universal features in financial time series have been discovered over the years, starting with Mandelbrot's discovery on commodity price patterns [1] in 1963. In the recent years some new universal patterns related to scaling properties have been discovered [5–7], and have been confirmed across different financial markets [8–11].

Before presenting the various stylized facts of financial time series, some basic quantities are defined first.

Asset price S_t

The price of an asset at time t can refer to the transaction price of a stock of a company like IBM, or a commodity like cotton or gold. Exchange rate can also be considered as a price and it shares many similar stylized facts with stocks and commodities. Some studies use slightly different definitions of price (as in some cases [12], quote price is incorporated to analyze price changes), but we stick to the most conventional definition of transaction price.

Return r_t

When specifying return, a time interval τ must be specified as a parameter. It refers to the proportional changes in the prices before and after time τ. τ can be chosen to be from 1 minute to a few months. In this study we use log return as it is most frequently used in other literatures due to its additive nature, and under most circumstances (when price change is not too big) it is equivalent to raw returns. Let s_t denote the log value of price S_t, i.e. $s_t = \ln S_t$, the mathematical definition of return is

$$r_t = \ln(S_t/S_{t-\tau}) = s_t - s_{t-\tau} \tag{1.1}$$

Trading Volume V_t

Volume is defined as the total amount of asset being traded during the time interval τ. In terms of stocks it is simply the number of shares changed hands during that time. There are studies using traded value which is the total traded dollar value within a time interval τ. These two definitions are largely the same given no significant change in price. For each trade i within the time interval $t-\tau$ to t, the number of shares traded is q_i. Here we define the trading volume at time t is

$$V_t = \sum_{t-\tau<i\leq t} q_i \tag{1.2}$$

Sometimes the number of transactions is used instead of real time as time scale. There is merits for both choices of time scales, yet for simplicity we only use real time scale in our analysis.

Number of trades N_t

The number of trades at time t simply refers to the total number of transactions during a time interval τ.

1.1.1 Absence of autocorrelation in returns

It is widely accepted that prediction of price changes is almost impossible at a meaningful time scale. This is in line with efficient market hypothesis (EMH), as the market is efficient at arbitraging the correlations in returns when they appear, until no more correlations exist for market participants to exploit. Mathematically, the extend of predictability in stock market can be measured by autocorrelation function (ACF) of return r_t:

$$\varrho_l = \frac{\langle (r_{t+l}-\mu)(r_t-\mu)\rangle}{\sigma^2}, \tag{1.3}$$

where $\langle \cdots\cdots \rangle$ refers to the average, μ is the mean value of returns $\mu = \langle r_t \rangle$, and σ^2 is the variance of reutnrs $\sigma^2 = \langle (r_t-\mu)^2 \rangle$. If price at time t is uncorrelated with price at a later time $t + l$ for any value of l, it means there is no predictability on returns, and $\varrho_l = 0$.

The fact that $\varrho_l \approx 0$ for any l larger than a few minutes has been verified by many different studies. Figure 1.1 gives a graphical representation of this empirical fact.

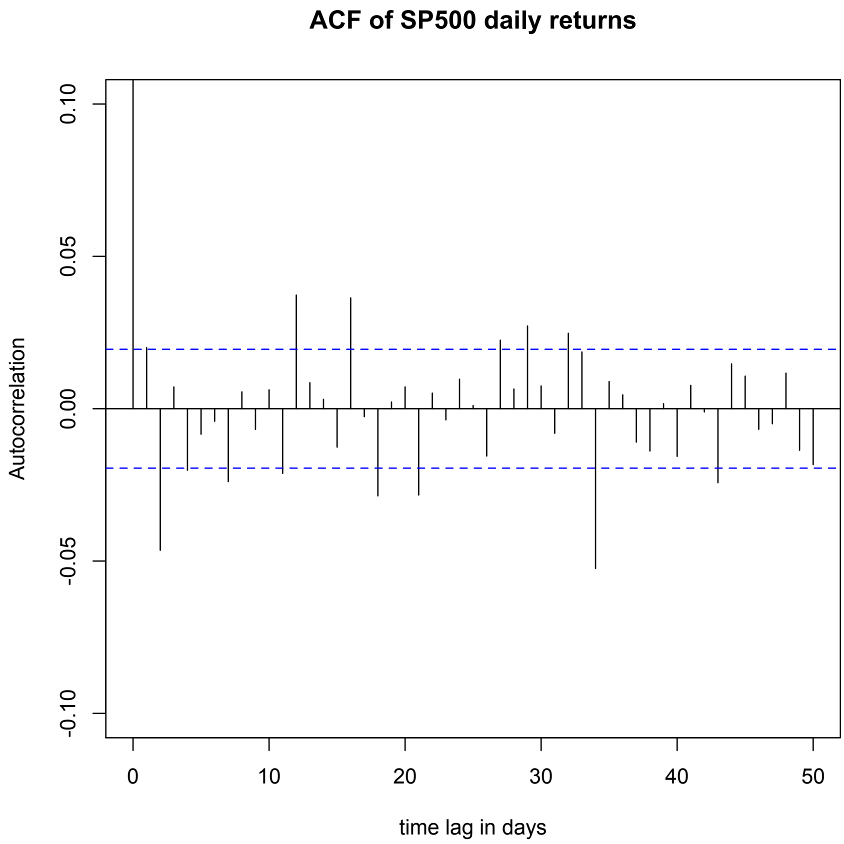


Figure 1.1: Autocorrelation of S&P500 daily price returns. The dotted line marks the noise level. The ACF of all lags are around noise level, signifying a lack of significant memory in daily returns.

1.1.2 Power-law probability distribution of large returns

The probability distribution of returns has received immense attention by academics and practitioners alike, as it gives the most straight forward representation of market fluctuations. Discovery of fat tail or excess kurtosis in return distributions can be tracked back to 1963, again in Mandelbrot's paper [1]. In simple words, fat tail means there is a significantly higher probability of large returns than Gaussian distribution. It was later accepted that return distribution follows a power-law at the tail region, meaning a straight line on log-log plot. Mathematically speaking, this means the cumulative distribution function has the form of $P(r > x) \sim x^{-\xi_r}$ when x is big – typically a few times the size of standard deviation σ. Here ξ_r is called the *tail exponent* of return distributions. The fact that whether the value of ξ_r has a universal value across different financial markets is still being debated, but it has been well accepted that its value is larger than 2, which means the distribution is outside Levy stable distribution as hypothesized in earlier literatures [5]. A tail exponent larger than 2 also means the variance - second moment - is finite. Some results indicate that the fourth moment of return statistics is finite, inferring a value of ξ_r larger than 4, yet others [13] show the contrary. Certain studies [14] argue that there might not be any universal value for ξ_r as it differs from asset to asset. The differences in opinion is possibly due to different methods in estimating the tail exponents, as well as the range of empirical data used in analysis.

The power-law tail has been repeatedly found in different markets during different periods on stocks or stock indices. When the sample size is large - in the order of millions - the tail exponents ξ_r of high frequency data has often been found to be close to 3 [2]. Similar results have been found in other markets including France, UK [8, 9], Japan, [10] and India [11], but criticism [15] were directed at such results for their lack of robust estimators used in evaluating the tail exponents. Such criticism arised as physicists mostly used graphical fitting to conclude the power-law tail nature of statistics, without rigorous statistical testing.

A reliable method for statistically estimating the tail exponent ξ_r is to use Hill estimator and its various implementations. We will discuss this method in detail in Chapter 2 as it is the one we use to carry out empirical analysis. While this thesis does not assert any universal value on the tail exponent ξ_r, empirically retrieved ξ_r values are used to verify our models and theories, and $\xi_r \approx 4$ has been found.

1.1.3 Long memory of absolute and squared returns

Despite the apparently erratic nature of price fluctuations, the size of price fluctuations tend to remember their own pasts. In other words, big price changes (without regards to the sign of the changes) tend to be followed by big changes and small changes tend to be followed by small changes. Mathematically speaking, it means the auto-correlation of absolute returns (or squared returns) decays slowly with time lag l, and remains significantly above zero with l value up to a few weeks or even months.

The autocorrelation of absolute returns is defined similarly to 1.3:

$$\rho_l = \frac{\langle (|r_{t+l}| - \mu_{|r|})(|r_t| - \mu_{|r|}) \rangle}{\sigma^2_{|r|}}, \tag{1.4}$$

with $\mu_{|r|} = \langle |r_t| \rangle$ as the mean of absolute returns, and $\sigma^2_{|r|} = \langle (|r_t| - \mu_{|r|})^2 \rangle$ is the variance of absolute returns.

The autocorrelation of squared returns can be defined in the similar way:

$$\rho'_l = \frac{\langle (r^2_{t+l} - \mu_{r^2})(r^2_t - \mu_{r^2}) \rangle}{\sigma^2_{r^2}}, \tag{1.5}$$

with $\mu_{r^2} = \langle r^2_t \rangle$ as the mean of squared returns, and $\sigma^2_{r^2} = \langle (r^2_t - \mu_{r^2})^2 \rangle$ is the variance of squared returns.

The slow decay of autocorrelation with increasing time lag l has been well documented with high frequency returns as well [16–18], with the value of l being on the scale of seconds to minutes. It must be noted that the existence of autocorrelation in absolute returns does not contradict with that of returns, as the earlier does not involve the sign of r_t. Figure 1.2 provides an graphic representation of this fact. This means while it is hard to profit from predicting price returns, it is possible to predict with some precision on the size of price fluctuations. This leads one to suspect that the lack of autocorrelation in return itself is due to the lack of memory in the sign of returns, which is related to the memory in trade signs ($+$ for buy trades and $-$ for sell trades). As we will see in the following section this is actually not the case - trade signs do have some autocorrelations.

One additional remark has to be made here regarding the use of autocorrelation

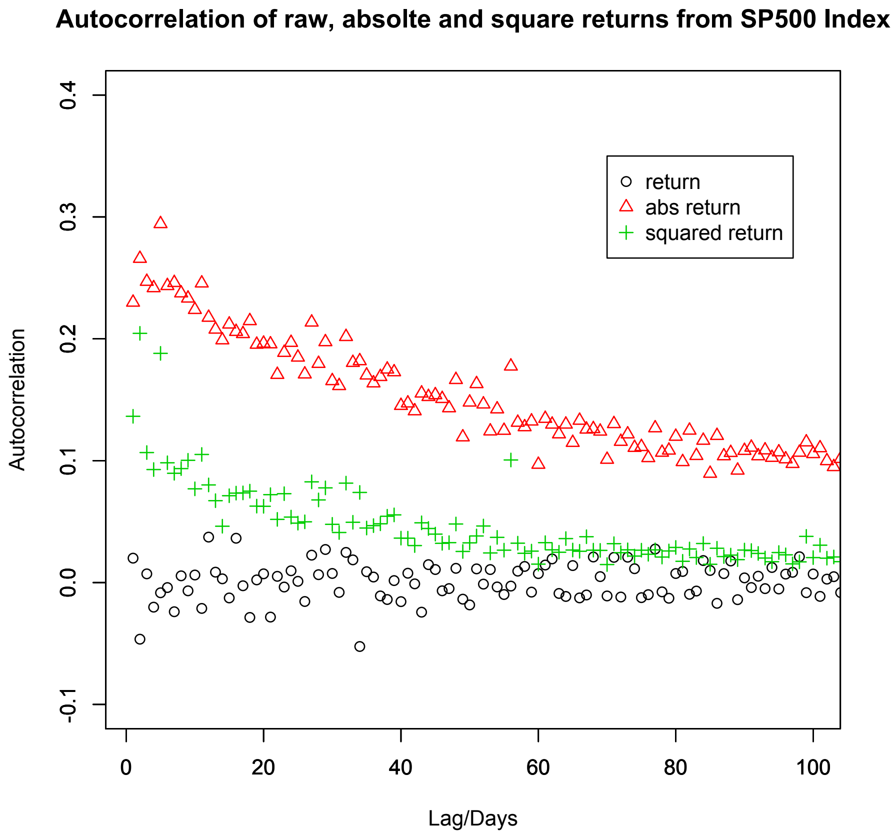


Figure 1.2: Autocorrelations of raw returns, absolute returns and squared returns from S&P500 daily prices. While raw returns do not show significant ACF, both absolute and squared returns show the opposite. Absolute returns have larger ACF than squared returns, meaning more persistent memory.

as a statistical measure. The reliability of autocorrelation in financial time series has been questioned [19] due to the existence of fourth moment of returns. As the measure autocorrelation of squared returns is directly related to the fourth moment, and there is an ongoing debate on whether the fourth moment is finite (Section 1.1.2), its validity could be questionable. Since this is a matter of mathematical interest, this study does not discuss in depth on this issue.

1.1.4 Memory in trade signs

Trades signs are defined as the buy/sell nature of a transaction. It has to be emphasized that while every transaction involves a buyer and a seller simultaneously, it is usually initiated by one of them - either the buyer or seller. There are usually two types of traders in the market in terms of order placement - market takers who place market orders to buy/sell, and the transaction takes place at the best price; market makers who place limit orders that are taken up by market orders passively. The job of market makers is to provide liquidity to the market so that a buyer does not need to wait to find a seller to have a transaction. Hence if a trade is initiated by a buy/sell market order, the sign of the trade is assigned as +1/-1.

It has been observed that trade sign has a long memory as well [20–22]. One explanation lies in the behaviors of large block of trades by institutional players. In order to avoid market impact, large trades needs to be broken down and executed in smaller blocks over days or weeks, resulting in correlated trade signs. In this case, if the market makers place limit order price randomly, they would suffer a loss due to the one-direction movement of price; thus market makers tend to adjust the bid/ask price accordingly to minimize such a price trend. In [23], these behaviors are investigated and used to explain the absence of autocorrelation in return despite a clear autocorrelation of trade signs. In [24], the detailed strategy of breakdown of large chunks is used to reproduce the various distributional properties in return, volume and number of trades.

1.1.5 Power-law distribution of trading volume, size and number of trades

Compared to the abundance of literature on return statistics, trade volume distribution receives much less attention. Although trading volume and number do not reflect the price dynamics directly, they are very useful for deciphering behavioral patterns of market participants.

[25] pointed out that the distributions of trading volume in both real time and trade time scales has an asymptotic power-law tail with exponent less than 2. This fact was later verified in UK and France stock market [3] with similar conclusion on the tail exponents. Several estimators including Hill estimator and MS estimator [26] give similar tail exponents below 2 as well. The same study also finds the distribution of trade sizes has a power-law tail with exponent less than 2.

Similar to trade volume and size distribution, the number of trade distribution has also been examined. [27] has found that its distribution also has a power-law tail, with exponents between 3 and 4. [9] has confirmed this fact for both US and France market.

Overall, power-law tail has been observed across different quantities in financial market - price change, trade size, volume and number of trades. [24] gave a theory to unite all these power-law tail exponents, and the mechanism is based on the behavior of institutional players. While [24] has explained the various tail exponents, the model has not addressed the long memory property of price fluctuations.

1.1.6 Other stylized facts

Aggregate normality of price returns

It has been observed that if one increases the return time interval τ, the tail of return distribution gets thinner and eventually the distribution converges to a Gaussian (Normal) distribution [28, 29]. This means the distribution converges to Gaussian under aggregation of short time returns. But it has to be noted that this convergence to Gaussian distribution is very slow - even when τ goes to the scale of weeks or months the distribution still has a fatter tail than Gaussian.

Leverage effect

In financial market, there is a correlation between price returns and future volatilities. Bad news which induces price drops tend to increase future volatility, yet good news tend to decrease future volatility. This phenomenon has usually been associated with psychology of market participants, and very often been incorporated into different models by arbitrarily introducing a term that is asymmetric with respect to the sign of past returns. Since it is not one of the concerns in this work, the details of this effect is ignored.

Time reversal asymmetry

As time evolves only in one direction, reversing the time may result in different statistical properties. This apparently trivial fact is far from trivial as many time series models do not exhibits such properties, yet empirical findings shows the opposite is true [30]. This suggests that past squared returns influence future volatility in a different way of what is vice versa. In general, any model from ARCH family would have time reversal asymmetry, but their extent of asymmetry is very different from empirical values [31].

Order book properties

In recent years order book study has become a hot topic as there is a growing amount of order book data available. The order books are examined in detail to study market microstructure as well as trading behaviors at a tick-by-tick level [32–36]. From some of these studies, certain power-law distributions are also found with different exponents. While such studies are very important to understand market behavior at a high frequency times scale, it is not closely related to this work. Hence the findings are not presented here.

Among all the above mentioned properties of stylized facts in financial time series, fat tail for returns and long memory in volatility are the most prominent. Most studies has been focused on unraveling the possible underlying mechanism for these two properties, and various explanation has been given with some success. In the following section, we will present some agent-based models which could capture these properties with some success. We will also discuss the shortcomings of those models and their implications.

1.2 Agent-based models

The efficient market hypothesis (EMH) implicitly assumes the representative agent model, in which all agents behave in the same way and maximize the same utility function. On the other hand, the various stylized facts are not able to be explained by the theory. In recent years, behavioral economics has emerged to counter EMH for its 'rationality' assumption, as empirical evidence has shown a range of irrationalities among people, not to mention the most basic ones like 'panic' and 'herding'. On a separate path, while various statistical models, in particular GARCH ([37]) and its related models, are successful in capturing some of these features, they are not able to convey a good understanding of the underlying trading behaviors. The urge to understand market dynamics and have better models to reproduce price fluctuations calls for a new approach of modeling – agent-based models (ABM), in which irrationality is assumed, and interaction among agents are incorporated. Heterogeneous behaviors and interactions of ABM mean it is much more complex than the traditional representative agent models, and many of them could only be solved with the help of computer simulations yet analytical result can only be obtained for the most simplistic cases. Literature on irrational agent behaviors emerged in the past decades, and many studies have taken these findings into ABM – herding, myopic interest, or threshold trading, just to name a few. Other than understanding market dynamics, ABM could also help policy makers to regulate the market or define better market structure of trading protocols [38, 39].

Although there are many physicists working in agent-based models, this stream of work started in the home ground of economics back in 1950, when William Philips – who discovered Philips Curve – used a hydraulic machein in his simulation of macroeconomics [40]. [41] is one of the first agent-based models on modern financial market, and the work was collaborated between economists and a physicist. It produced the cyclic behavior of financial market, yet most of the stylized facts were not discussed. Lux-Marchesi Model [29] introduced the idea of fundamentalist-chartist interaction into their model and successfully reproduced power-law tail. This idea of fundamentalist-chartist was later well accepted by many ABM models as a grounding assumption. Herding behavior was also incorporated into random interaction among traders by Cont and Bouchaud [42] to produce power-law tail in returns, and this model was modified further by many to produce other stylized facts. The effect of updating decision threshold has been studied by Cont [43] trying to explain the phenomenon of volatility clustering. There have been various models focusing on agent behaviors of order books rather than price itself. One recent study by Farmer [44] provides a very detailed empirical study on the order book dynamics, and successfully reproduced price dynamics on certain stocks. All these models have their strengths and weaknesses, and will be discussed in the following sections.

1.2.1 Heterogeneous agent-based model (1999)

This model proposed in [29] was one of the first in reproducing fat tails with volatility clustering. Rather than simply assuming all market participants are rational (as what is stereotyped as 'fundamentalists'), it assumes another group of agents who do not respect fundamental value in their decision making.

The market participants are classified into two groups: fundamentalists who believe the stock price will get back to fundamental value in the future, and chartists who either think the price will go up - optimists - or will go down - pessimists. All of the traders (agents) update their strategies (fundamentalists, optimists, pessimists) at every time step, which means they may change their trading strategies at any time with certain probabilities. The mechanism that induces this strategy switching is governed by some mathematical formulas and a set of parameters which are complicated, yet the underlying rationale was simple: there are two motivations for any agent to change strategy – profit and herding. 'Profit driven' means if another strategy has made more profit in the previous time step, the agent is more likely to change to that strategy. 'Herding' means if there are more agents adopting another strategy, one would be more likely to change to the strategy.

At each time step, there is a fundamental price S_f which is randomly generated from a normal distribution, and the market price is determined by excess demand of all agents – the difference between the number of buyers and sellers.

The set-up of the model has ensured the market behaves around a equilibrium state [45], where there is a balanced population in both groups. Once the price deviates from fundamental value by too much, more chartists will switch to fundamentalists and stabilize the market back to this equilibrium. This resembles a system fluctuating at a critical state, in support of the hypothesis market by econophysicists that market is at a critical state similar to phase transitions. Power-law tail exponent is found to be within 1.9 to 4.6, overlapping with the range of empirical

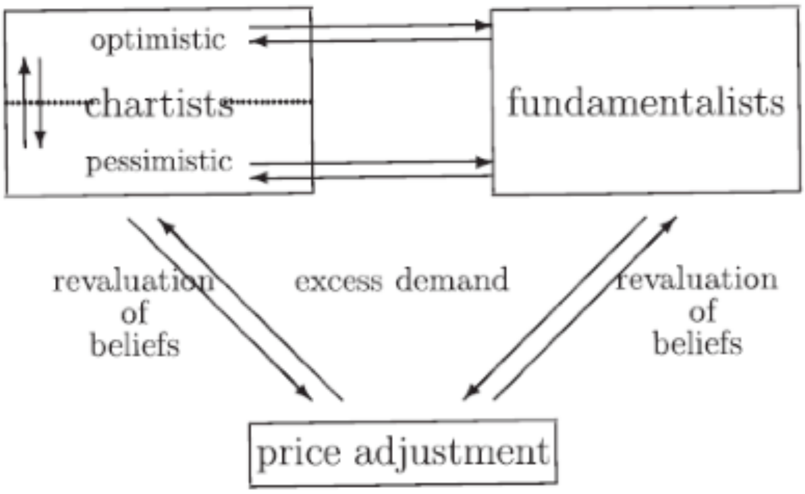


Figure 1.3: Illustration of the mechanism behind the heterogeneous agent based model [29]. The model assumes three types of strategies used by market participants – optimistic and pessimistic strategies, as well as fundamentalists. The traders switch from one strategy to another from time to time depending on the performance of each strategy and the number of people using each strategy.

findings, and volatility clustering has also been produced by the model to some extend.

As the first model to quantitatively produce stylized facts, this model has been very successful, and follow-up studies has been carried out to examine the different variations of this model [46, 47]. On the other hand, it has the drawbacks of being complicated and involves too many parameters. Furthermore, [46] has pointed out that the desirable stylized facts vanishes when the number of agents is larger than a few thousands – in real market this number is much larger. Only if the strength of herding effect grows with agent population, the stylized facts could be simulated. But this idea of increasing interaction in herding is hard to justify in real market. One possible way to get around this limitation is to introduce interaction network structure as what has been done in [47]. In the original model, it was assumed every agent can interact with everyone else in the herding mechanism. Yet in [47], it was assumed that each agent can only herd within the people he/she knows or interact with. The interaction structure is defined in terms of network structure of agents, with each individual agent being a node, and each pair of interaction as a link. With such modifications, the model may still suffer from the 'size effect' unless very specific interaction structure is used.

1.2.2 Herding (percolation) model (2000)

Percolation is a term used in graph theory. It refers to the incidence when a connected path exists through the whole graph (sometimes referred to 'network'). It has been studied both in physics and chemistry, where a material can be thought of as a graph in limited dimensions - 2 or 3 dimensions. When borrowed to the theory of networks, it also refers to the incidence when a giant connected cluster exist in a network, and many critical phenomena appears at percolation.

To illustrate percolation in a network, let us first look at the case for a random network. For a system of N lattice points (nodes) in D dimension, p is defined to be the probability that two randomly chosen nodes have a link between them. When a set of nodes with total number C is connected, meaning there is a path of links connecting any two nodes in this set, a cluster of size C is formed. For a given value of p, the distribution of cluster sizes C can be determined. When the value of p is at a critical value p_c, a giant cluster will be formed with the size comparable to N, and this phenomenon is called percolation. At percolation, the cluster size distribution will follow a power-law depending on the dimensionality of the system. A schematic illustration is shown in Figure 1.4.

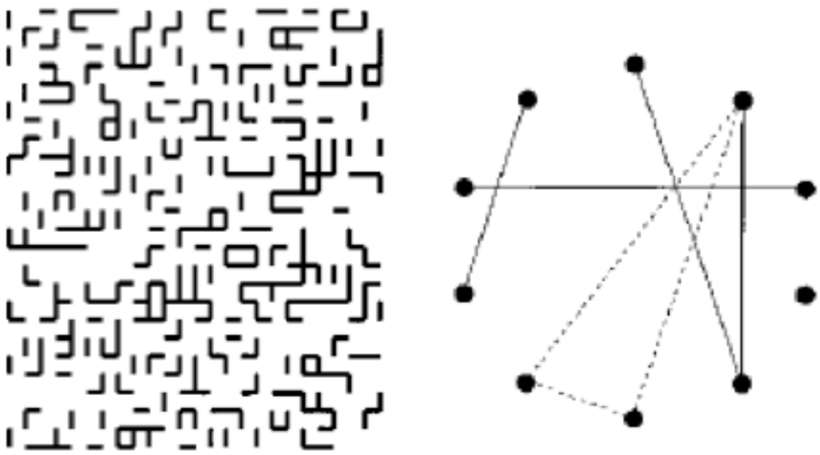


Figure 1.4: Illustration of percolation in networks of different dimensions. The figure on the left shows percolation on a two dimension square grid system. The right figure shows the percolation on a random network, i.e. the links are randomly generated to connect the nodes on the graph. Adopted from [42]

In the model developed by physicists [42], every agent can be thought of as one node, and a link in the network can be interaction between to nodes, as not all agents know every other agent in the market. In each cluster, all the agents share the same opinion, being it buy, sell or do nothing. In other words, the cluster of agents act unanimously. At each time step, cluster i will make a decision $\Psi_i(t)$ to buy/sell with probability a, and hold with probability $1 - 2a$. Mathematically it is defined as

$$P(\Psi_i(t) = +1) = P(\Psi_i(t) = -1) = a \tag{1.6}$$

$$P(\Psi_i(t) = 0) = 1 - 2a \tag{1.7}$$

Hence the total demand in the market is given by

$$ED_t = \sum \Psi_i(t), \tag{1.8}$$

and the price change at each time step is assumed to be proportional to ED_t,

$$r_t \propto ED_t \tag{1.9}$$

With this construction, it can be shown that power-law tail CDF of returns can be produced if $p \approx p_c$. Though being simple in its construction, the tail exponents predicted by the model is not close the the empirical values, and the lack of higher moment autocorrelation is also at contrary to real data. Various modifications have been added to the model including trend following behavior, increasing market activity when price change is big [48], and some interesting results appeared. Ising-like modifications [49] were also added and had good results. Similar models using magnetic spin interactions [50, 51] were also proposed with good results.

Despite the fact that the model is able to produce certain stylized facts, its recognition outside physics community is not so well received. It has been criticized for being too 'physical' and lack economic intuition as why people behave in a lattice structure like materials. What was also being questioned is the value p being around the critical value of percolation. There is no argument or evidence on why people should behave at this critical state. The power-law tail exponent in return distribution is also far from the empirical values unless further assumptions are made. Nevertheless, the model provides good insights on the possible mechanism of certain stylized facts.

1.2.3 Threshold updating model (2007)

Unlike the previous models, [43] aims primarily at explaining volatility clustering. In the model, each agent is assumed to react to public information at each time interval. The public information is compared with their own decision threshold to enact a buy/sell transaction. Instead of assuming a homogeneous reaction to this exogenous information, the thresholds of agents are randomly distributed, and every once in a while this threshold value is updated with certain probability. This behavior of updating threshold was deemed as a phenomenon of investor inertia, meaning each agent will used the same strategy/threshold for a considerable duration before switching to another. It is argued that this simple behavior reflects the regime switching mechanism proposed in [52] that generates volatility clustering. Actually region switching in the context of market could be viewed as the economic version of phase transition behavior in statistical mechanics.

In the model, there are N agents each having a trading threshold $\theta_i(t)$ participating in the market. ϵ_t is the common signal (public information) received by all agents, and it is assumed to follow normal distribution $\epsilon_t \sim N(0, D^2)$. ϵ_t represents the common signal received by all of the agents. At every time interval each agent compares the signal ϵ_t with his/her own threshold $\theta_i(t)$. If $|\theta_i(t)| < \epsilon_t$, agent i will not trade. If $\theta_i(t) < -\epsilon_t$, the agent will sell, i.e. $\psi_i(t) = -1$, and if $\theta_i(t) < \epsilon_t$, the agent will buy, i.e. $\psi_i(t) = 1$. Return r_t at each time step is calculated in the same way as section 1.4 – based on aggregate demands. At each time step, every agent has a chance to update the trading threshold $\theta_i(t)$ to the most recent volatility $|r_t|$, as supported by the empirical study [53]. The probability of switching to the new threshold is $1/q$.

As a result, the model has simple mechanisms and permits realistic estimation of certain parameters by fixing the trading time steps at daily scale. The simulation results indeed shows a slow decaying autocorrelation of absolute returns, and excess volatility compared to the volatility of public information D. Uncorrelated returns were also observed and analytically proven. On the other hand, power-law tail in return distribution is not observed. Considering the model primarily focused on explaining volatility clustering, it has given a explicit account that long memory in volatility is due to the frequency that agents update their trading threshold $\theta_i(t)$. Mathematically, the length of the memory depends on the value of q.

1.2.4 Model of order book dynamics (2008)

Modern trading platform has seen a transformation more than a decade ago, when all transaction details started to be recorded electronically. As discussed in section 1.1.4, a transaction occurs when a market order hits a limit order. There are usually a lot more limit orders than market orders in the trading platform, and both types of orders are recorded in ‘order book’. [44] has provided a detailed study on order book dynamics - both order placement and cancellation. By fitting the statistics of order book dynamics to different functional forms and parameters, the authors constructed a zero-intelligence model capable of reproducing stylized facts in price returns. The model is constructed in the following steps:

1. Simulation of order placement

   (a) Generate order sign (fractional Brownian Motion)

   (b) Generate order position (Student t distribution)

2. Simulation of order cancellation

   (a) Dependence on order position (exponential)

   (b) Dependence on order imbalance (linear)

   (c) Dependence on total number of orders (inversely proportional)

With all the characterizations of order book dynamics in place, the model simulates price dynamics for 24 different stocks, and several stylized facts on price have been recovered with numerical accuracy. The statistics of bid-ask spread have also been reproduced.

Power-law tail of return distribution and volatility clustering were observed, and the value of tail exponent ξ_r can be reproduced with numerical accuracy for half of the stocks analyzed. In conclusion, the authors pointed out that stylized facts may just be a trivial and inevitable outcome of order book dynamics, and studies on behaviors of traders could be focused on order dynamics rather than price dynamics.

As a zero-intelligence model, this model suffers from several drawbacks. First is the predictability of returns in the simulation result, or serial correlation in returns. The model assumed correlated order signs from empirical evidence, but it neglects the adjustment of liquidity by market makers [21] which prevents correlated returns [54]. One other drawback of the model is that the simulated volatility has shorter memory than empirical findings, and this could also be attributed by certain agent behavior that is ignored in this zero-intelligence model. With all these being said, it could be possible that the lack of any behavioral ingredients in the model makes it an insufficient model of price dynamics, yet it could be a good starting point for order book studies.

1.3 Summary on existing literature

There are numerous other agent-based models on financial market, some of them are not discussed here but are very important as well. The famous Santa Fe Institute artificial stock market model [55] and minority game [56] are two of them. There are other literatures on evolutionary models [55, 57, 58] which share certain features with agent-based models as well. Such models have the gist of evolutionary theory in biology, and produce some stylized facts and other interesting observations. On the other hand, calibration of such models is almost impossible, thus direct comparison with empirical data is not meaningful [43].

What is common to all of those studies is they investigate financial economics using a bottom-up approach, and progress has been made in various aspects:

- While efficient market hypothesis fails, stylized facts can be reproduced by agent-based models, though the validity of their underlying mechanism is yet to be confirmed to gain wider acceptance.
- Agents have specific behaviors other than rationality. This is reflected by the failure of various zero-intelligence models which even dig as deep as into the

detailed dynamics or order books.

- The financial market is not always at equilibrium - even if such an equilibrium exists. It is the fluctuation between different regimes - in terms of trading strategies or thresholds - that results in various stylized facts.
- Most of the existing models suffer from the population effect. When the total number of agents becomes large, simulation results converge to normal distribution with no power-law tail. This problem can be got around if the interaction strength and structure are adjusted accordingly. In particular, the interaction strength among agents needs to increase together with agent population N. But a sound explanation for this needed.

The last point has a significant implication on the collective agent behaviors: in order for interaction parameters to grow with N, their interactions cannot be local, otherwise every agent has only limited number of interactions possible, and this interaction does not scale with N; if the interaction is global similar to the one in the threshold updating model in section 1.2.3, the interaction strength would naturally grow with population size N. The question is, what could be such global interaction mechanism? One possibility is the exogenous public information available to all agents. But this information happens at a frequency much lower than daily yet the various stylized facts exists at daily or even intraday level, and numerous studies have shown that the volatility of price fluctuations are much larger than what is perceived from public information (this phenomenon is also know as 'excess volatility'). One possible interaction mechanism is price itself, as every agent observes price change and acts on it. But exactly in what ways does price affect agents' strategies and trading behaviors? We will investigate this mechanism in chapter 3 from various empirical studies and theoretical motivations. Before going into the analysis of market mechanism, we present the data and methodologies in the following chapter.

Chapter 2

Data and Methods

The stylized facts exist in most of the developed market for a broad range of financial products including stocks, foreign exchange rates, commodities etc. In this study the subject of interest is US stock market, for it is arguably the most representative. The major sources of data are Wharton Research Data Service (WRDS) and *Yahoo!* Finance. WRDS is a database provided by The Wharton School at University of Pennsylvania for academic research purposes. It consolidates a broad range of financial databases including COMPUSTAT, CRSP, NYSE-TAQ and more [59]. The range of data is discussed followed by some methodologies used in the work, including the programming languages.

2.1 Data

Since this study focuses on US stocks, the most representative set of stocks from US stock market is chosen. Two sets of stocks are chosen: the 500 companies in the Standard & Poor's 500 (S&P500) index components and the 30 stocks in Dow Jones Industrial Average Index (DJI) components. The two sets of stocks are used in different stages of the work for different purposes.

Standard & Poor's 500 Index Components

The components of S&P500 index change from time to time, and the 500 companies chosen are the ones on S&P500 list as of July 2012. S&P500 index components represent the major companies in US stock market both in terms of their large market capitalization and high trading activity. They serve as a barometer for the whole US stock market and economy.

Since the financial market has undergone major transformations in the last half a century due to computerization in late 90s and the adoption of high-frequency trading in the recent decade, there have been structural changes in trading volumes. Hence the period of data studied is from 1997 to 2006, during when high frequency trading has not been massively adopted and program trading has already become a norm. During the ten-year period 1997-2006, the list of 500 companies in the S&P500 index components has been constantly updated. For the consistency of the study, the ones that have inactivity in trading during certain months of the 10 year period are taken out of analysis, so that every stock has a complete time series data. Overall, there are 309 stocks in the selection.

Dow Jones Industrial Average Index Components

Dow Jones Industrial Average Index has a much longer history than S&P500, with more than 100 years of data values available. It comprises 30 companies that are most representative of the different industrial sectors of US economy. Only part of the study uses the companies in DJI index, and the period for the data is from 1971 to 2010.

2.1.1 Stock holdings by Institutional owners

For each listed company in US stock market, *Yahoo!* Finance [60] provides the information on the major holders of the stocks, as well as the percentage held by institutional owners and mutual funds. For simplicity, all these investors are called institutional owners or fundamentalists throughout the study. The institutional owners are the ones who focus on the long term performance of the companies they bought, and have a strong emphasis on the fundamental values.

Figure 2.1 shows the percentage of holdings by fundamentalists in each of the 309 stocks. It can be seen that for most of the stocks, more than 60% of the outstanding stocks are held by institutional owners, with a considerable number of them being almost 100% held by institutional owners. The average percentage value in figure 2.1 is 82.9%. Since institutional owners hold their positions for long term investment, this means more than four fifth of the stocks are not changing hands frequently.

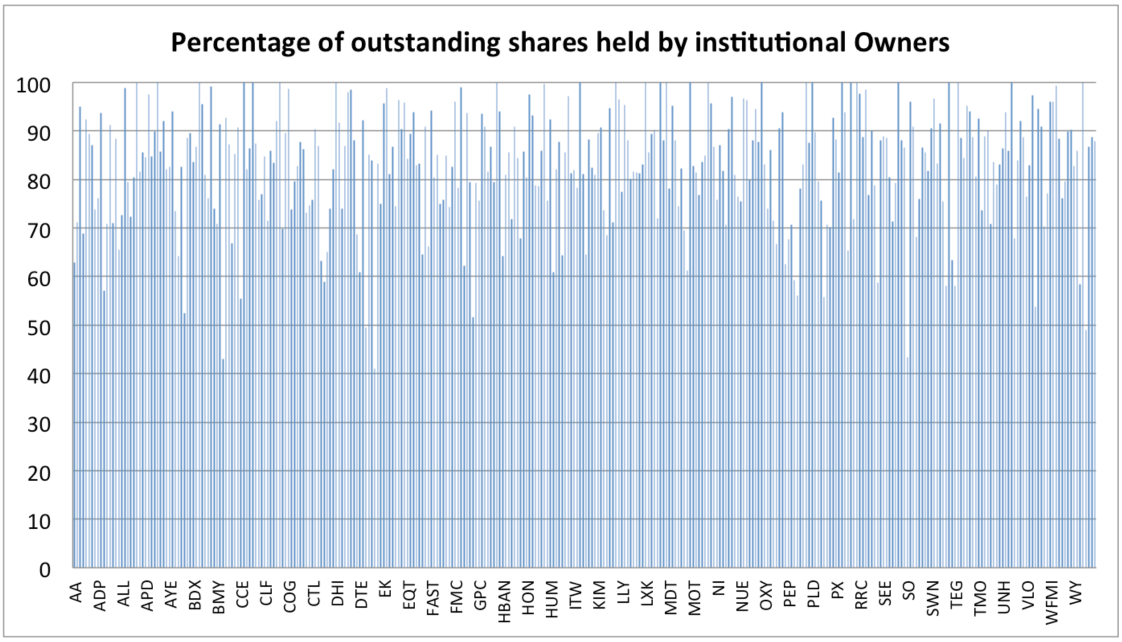


Figure 2.1: Percentage of outstanding shares held by institutional owners for the 309 stocks in S&P500 index components. A majority stake is held by institutional owners who invest for long term profits. The average of outstanding stocks held by fundamentalists is 83%.

2.1.2 Transaction data of US stocks

High frequency transaction data is retrieved from Trade and Quote (TAQ) database in WRDS. TAQ provides the trade details for every single trade including trade price, size, time. Price return r_t can be calculated based on equation 1.1 for different values of return intervals τ. There are 6.5 trading hours per day, hence the return interval τ can be taken at $1, 5, 10, 20, 60, 120, 360$ minutes within the same trading day. In every trading day, the price of the first transaction is the opening price, and the price of the last transaction corresponds to the closing price. High frequency returns are calculated based on equation 1.1 with different values of τ.

Low frequency data is obtained from COMPUSTAT database in WRDS. COMPUSTAT provides the total number of outstanding stocks, daily transaction volume, total number of trades, opening and closing price and other information. The opening and closing price at day t are labelled as $S_{t,open}$ and $S_{t,close}$. Hence the price return in the trading day is $r_{t,day} = \ln(S_{t,close}/S_{t,open})$, the price return outside the trading day (when there is no trading activities) is $r_{t,close} = \ln(S_{t,open}/S_{t-1,close})$, and daily return is simply $r_{t,daily} = \ln(S_{t,close}/S_{t-1,close}) = r_{t,day} + r_{t,close}$. For the simplicity of formality, r_t refers to $r_{t,daily}$ in this thesis.

2.2 Methodologies

2.2.1 Autoregressive conditional heterokedasticity models (ARCH)

In time series analysis, autoregressive conditional heterokedasticity (ARCH) models are widely used to simulate the empirical features of data. It was first developed in 1982 by R. F. Engle [4] to study the inflation of United Kingdom. It was later modified into many different variations in the study of financial time series.

ARCH model assumes that the variance of a time series changes from time to time, and this conditional variance depends on past data points. For a process with zero mean, at each time t, the random variable ϵ_t is given by

$$\epsilon_t = \sigma_t \eta_t \tag{2.1}$$

Here η_t is a random variable with zero mean and unit variance, and σ_t^2 is the conditional variance at time t modeled by

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdot + \alpha_q \epsilon_{t-q}^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 \tag{2.2}$$

In the case of financial time series, ϵ_t is the return time series after removing the mean, i.e. $\epsilon_t = r_t - \mu_r$. q characterizes the length of the memory in ϵ_t.

In its construction, ARCH model is able to capture the variation in variance/volatility of the time series, and the memory decays with length q. To better simulate financial time series, T. Bollerslev constructed a model called generalized autoregressive conditional heteroskedasticity (GARCH) [37]. It is a generalized model of ARCH in the sense that ϵ depends on the past volatilities σ_{t-i} in additional to past residuals ϵ_{t-i}. Mathematically, the conditional volatility is given by

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \tag{2.3}$$

When applying GARCH models to stock returns, larger p and q values can have better fitting but with poor predicting power, and most of the time the simplest case of $p = q = 1$ gives very good result. The problem is by fixing $p = q = 1$, estimation of parameters in equation 2.3 usually gives $\alpha_1 + \beta_1 \sim 1$. This let to the believe that the sum of the α and β coefficients is 1. Hence Integrated GARCH (IGARCH) model was proposed to simulate this effect by enforcing the relation

$$\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i = 1 \tag{2.4}$$

While GARCH and IGARCH could capture volatility clustering to some extend, they are not able to capture the long memory properties very well. In particular, the ACF of time series simulated by GARCH is decays too fast (exponential) compared with empirical data , yet the ACF from IGARCH is too persistent.

In order to simulate a more realistic memory effect, fractionally integrated GARCH (FIGARCH) model has been proposed [61]. FIGARCH model will be discussed later together with our own stochastic volatility model.

2.2.2 Tail exponent and Hill estimator

The convention for expressing cumulative distribution function (CDF) is to look at the probability that the random variable r is smaller than x, i.e. $P(r < x)$. For cumulative distribution with power-law tail, i.e. $P(r_t > x) \sim x^{\xi_r}$, it is more convenient to express the CDF in the opposite way: $P(r > x)$, so that the tail of the distribution would be a straight line on a log-log scale plot. Hence in this thesis, instead of plotting the CDF graphs by $P(r < x)$, the opposite convention of complimentary CDF $P(r > x)$ is used, and the plots are all done on log-log scale.

There are different methods to estimate the value of tail exponents ξ. Hill estimator is one of the most frequently used. It is a statistical estimator developed by Bruce M. Hill [62] for estimating power-law tail exponents. Its implementation is relatively simple. For a discrete time series data $x_1, x_2, x_3, \cdots, x_n$ with power-law tail distribution, it is first rearranged in decreasing order $y_1 \geq y_2 \geq y_3 \geq \cdots \geq y_n$. Since the power-law distribution usually occurs at the tail of the distribution, only the biggest values in the time series is used. In case the power-law distribution is valid for the largest k values, only y_1 to y_k is used for Hill estimator statistics:

$$\gamma_k = \frac{1}{k}\sum_{i=1}^{k}\ln(\frac{y_i}{y_{k+1}}) \tag{2.5}$$

The tail exponent is then calculated as the inverse of Hill estimator given by

$$\xi = \frac{1}{\gamma_k} \tag{2.6}$$

The difficulty of using Hill estimator is usually associated with the choice of threshold k. If k is too small the estimation has a large uncertainty; if it is too big then the region outside the power-law tail would be included and results in error. The accuracy of Hill estimator can be improved through resampling, i.e. bootstrap method [63] or jackknife method [64].

2.2.3 Monte Carlo simulation

The statistical properties of our models are investigated using both mathematical analysis and computer simulation. While mathematical analysis gives the fundamental functional form, but the complexity of models usually do not permit explicit analytical solutions. In such cases, computer simulation can be used to generate random samples such that the exact statistical properties of the model can be examined. In this work, Monte Carlo simulation is extensively used to generate datasets from the models, and the results are being compared to both analytical solutions and empirical data.

2.2.4 Programming languages

The whole research involved intensive data analysis and modeling, and different computer programming skills are suitable for each step.

For empirical data consolidation, the size of raw data from WRDS database is about 60 gigabytes, and different stocks are mingled together in the raw data files. In order to syphon through such a large data file and separate each stock into individual files with useful formats, UNIX shell programs are used. Shell programs are used in combination with stream editors *sed* and *awk*. The advantages of using stream editors *sed* and *awk* is that they can take large input files by reading them line by line, so that considerably smaller memory is needed from computer hardware. They are also simple to implement for the purpose of taking useful information from well-formatted data files. In the raw data from WRDS there are redundant transaction data and cancelled/modified transactions. UNIX stream editors are also good at cleaning such data with simple implementations.

Some of the cumulative distributions and Hill estimators are calculated using C/C++ for the mathematics involved. The agent-based model is also programmed in C/C++ as it provides flexibility of tweaking different parameters and realizations of various behaviors.

Monte Carlo simulation on our stochastic volatility model is done using R language. R is a system adapted for statistical computation and graphics, and widely used by statisticians and financial economists. It has ready-to-use packages for various statistical analysis including cumulative distribution and autocorrelation. It also has packages for many statistical models including autoregressive (AR), moving average (MA), ARCH, GARCH and its derivative models. In additional to such models, *R* provides packages of parameter estimation for the models.

Chapter 3

Empirical Data Analysis

Physics is a discipline where observations are made to construct theories, which are again verified by empirical analysis. In econophysics, the same path is being taken. Therefore before constructing any theories, let us begin by looking at the empirical stylized facts in detail.

Since we have the high frequency data with the resolution to every transaction, we can investigate the power-law tails of return distributions at different time scales, from every minute to daily. As most of the news event happens outside trading hours to avoiding a drastic market swings, it is worth investigating the differences between day and night returns as defined in section 2.1.2.

There are 309 stocks in our selection, and the data spans ten years from 1997 to 2006. In each year, there are approximately 250 trading days, so overall there are 772,500 daily data available, with approximately 2,500 for each stock. The power-law distribution for returns is observed for the top 5% of data or less, thus there is less than 75 data point available for each stock. This number is too small to have any accurate estimation of the tail exponents. Therefore, all the 309 stocks are aggregated into a single distribution in the analysis. Every stock has different values of mean and volatility in price returns. If the raw returns are aggregated, certain stocks may dominate the extremely large returns, and have uneven contribution to the tail part of the distribution. Hence each stock returns time series is normalized to zero mean and unit variance before putting together into a single distribution.

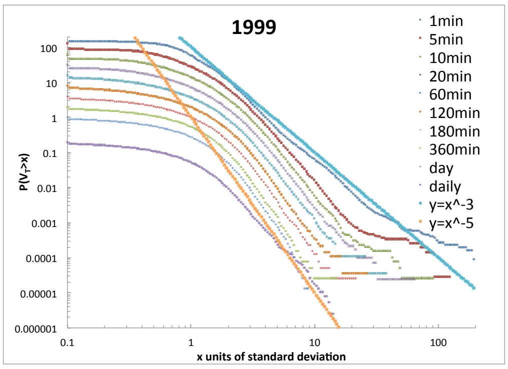
For example, a stock i has a return time series $r_{i,t}$ calculated from price by equation 1.1, a normalized time series is constructed:

$$\frac{r_{i,t} - \mu_i}{\sigma_i} \tag{3.1}$$

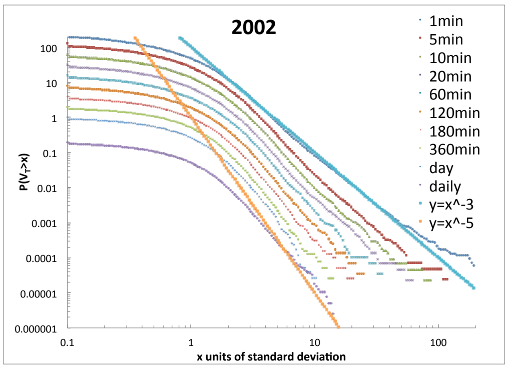
Here μ_i and σ_i are the mean and standard deviation of stock i.

3.1 Returns of different frequencies

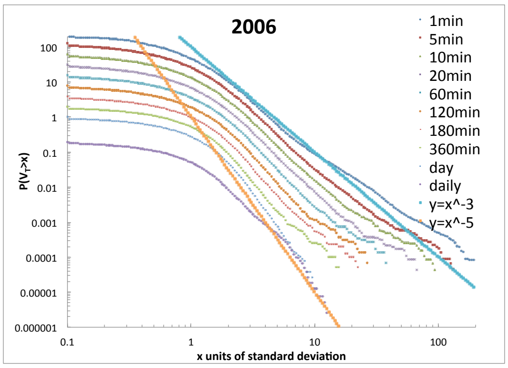
For high frequency returns, return time interval τ is chosen to be 1, 5, 10, 20, 60 120, 180 and 360 minutes. For each τ, there is a aggregate distribution. Thus we can investigate how the distribution changes with increasing return interval. Figure 3.1 presents the log-log plot of absolute return distribution with the different τ values. As argued in section 2.2.2, the cumulative distribution is plotted as $P(r > x)$ instead of $P(r < x)$. All of the distributions have a straight line at the tail region of their distributions, and the tail exponents are from -3 to -5. This is in line with pervious literature findings.



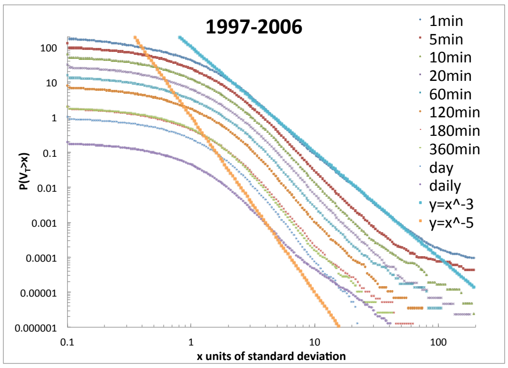
(a)



(b)



(c)



(d)

Figure 3.1: Cumulative distribution of high frequency returns. a-c shows the CDF of different years, and d shows the CDF of all 10 years aggregated together. The tail of CDF becomes fatter as the return frequency becomes higher, and tail exponents varies between -3 to -5 for all years.

Figure 3.1 a-c illustrate the distributions for three of the ten years studied. It is clear that the distribution at higher frequency has a fatter tail than the ones with lower frequencies, regardless of which year is being looked at. In some years, it seems even plausible that day return (at τ = 390min) has no power-law tail, whereas the daily returns still show a clear power-law tail behavior. Figure 3.1d shows the distribution for the whole ten-year period (normalization is done for every stock for the ten years rather than every one year). The tail exponent increases from 3.2 at τ=1min to 4.1 at τ=390min, and drops to 3.5 at daily level where overnight return is included. However, the tail for day returns calculated from 10-year period is conceivably smaller than individual years, showing that at long time scales, day return distribution has a fatter tail under aggregation of individual years, but it is still not fat enough to match daily returns (which has an additional overnight return component).

Figure 3.2 shows the values of tail exponents for different years at different return interval τ. For the estimation of tail exponents using Hill estimator, the threshold for tail region is 1%. It is obvious that the tail exponent increases slowly from a value close to 3.5 at τ=1min to a value close to 5 at τ=390min, and this trend is consistent for all years. For daily returns, the tail is fatter than day returns with a value of ξ_r around 4. The fatter tail is possibly due to the fact that overnight price changes are incorporated into daily return in addition to day returns. This phenomenon will be studied in detail in another section.

The thinning of tails with higher frequencies is called aggregate normality mentioned in section 1.1.6, which says that the distribution converges to normal distribution under aggregation, as low frequency returns are simply aggregations of high frequency ones. But it can be seen from figure 3.1 that this convergence is rather slow. From τ=1min to τ=390 min, 390 of 1min returns are aggregated to form a single τ=390min return, yet the tail of the distribution is still significantly bigger than normal distribution.

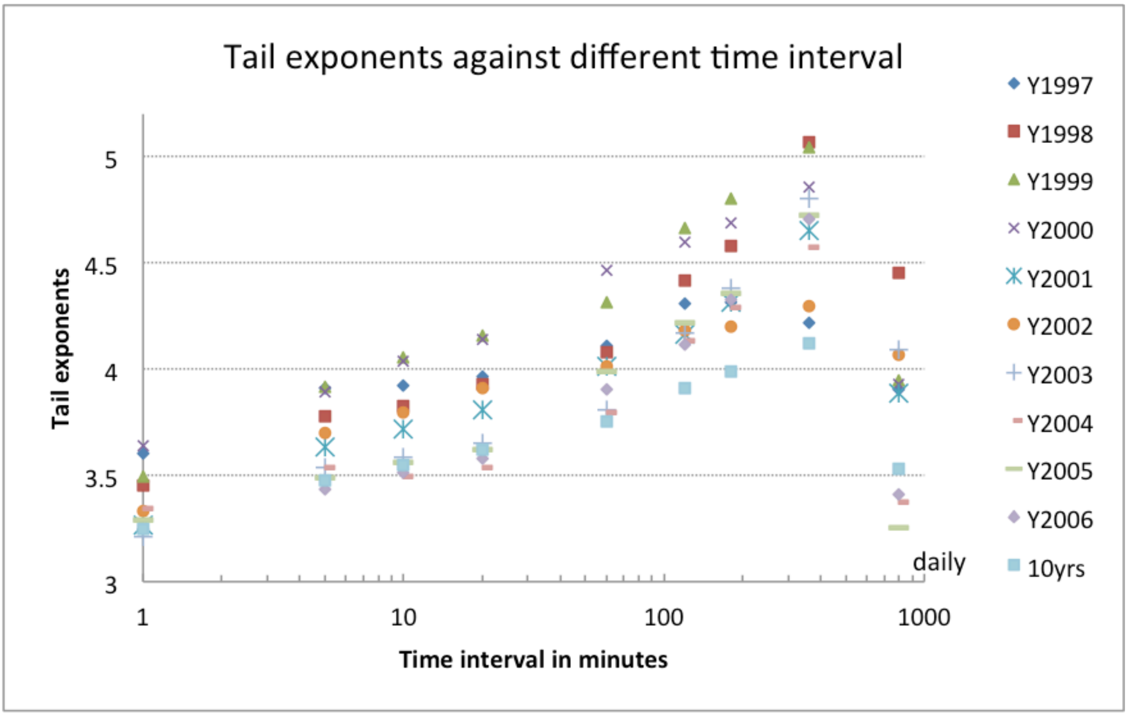


Figure 3.2: Tail exponents for distributions at different frequencies each year. There is a increasing trend in the value of ξ_{r_τ} as the return interval τ increases, yet the value is always between 3 to 5. At daily frequency there is a evident drop in the value of ξ_r from day returns of $\tau = 390$min.

3.2 Tail exponents of trade-by-trade returns

It is possible that the decreasing tail exponent value with lower return frequency is due to decreasing sample size. In order to investigate this, a trade-by-trade return r'_t is defined:

$$r'_t = \ln \frac{S_t^+}{S_t} \tag{3.2}$$

Here S_t^+ and S_t are the prices of two adjacent trades at time t, with S_t^+ happening right after S_t. For a particular value of τ, in order to have the same sample size as r_t time series, one value of r'_t is calculated after every τ minutes. For example, for every return interval of 5 minutes, r'_t is calculated at $t = 0, 5, 10, 15, \cdots, 385, 390$

minutes.

As shown in figure 3.3, for each year the tail exponents change very little upon increasing τ from 1min to 180min. Therefore it can be concluded that the decreasing tail exponent value ξ_r is not related to shrinking sample size nor not enough coverage of volatility variations.

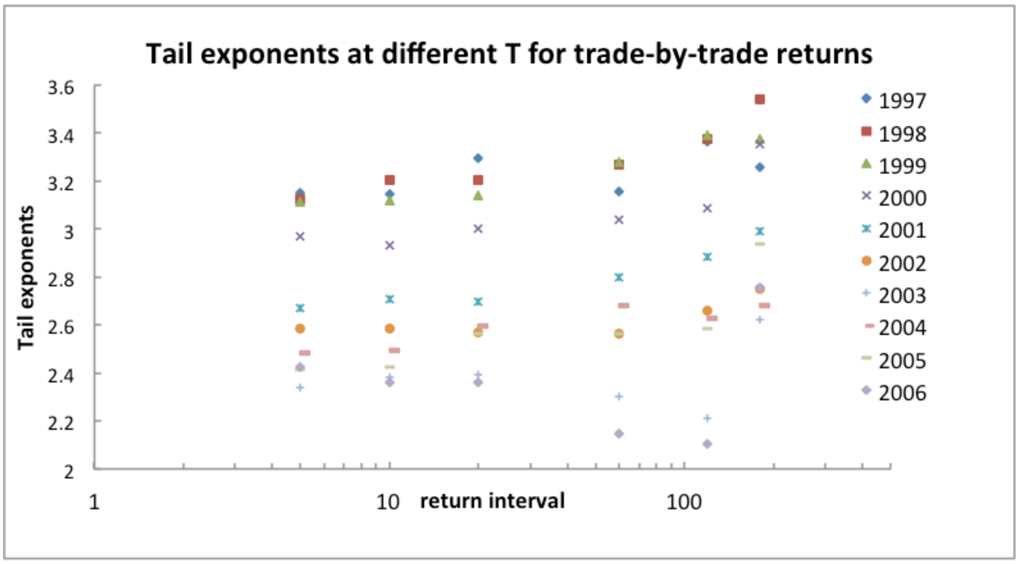
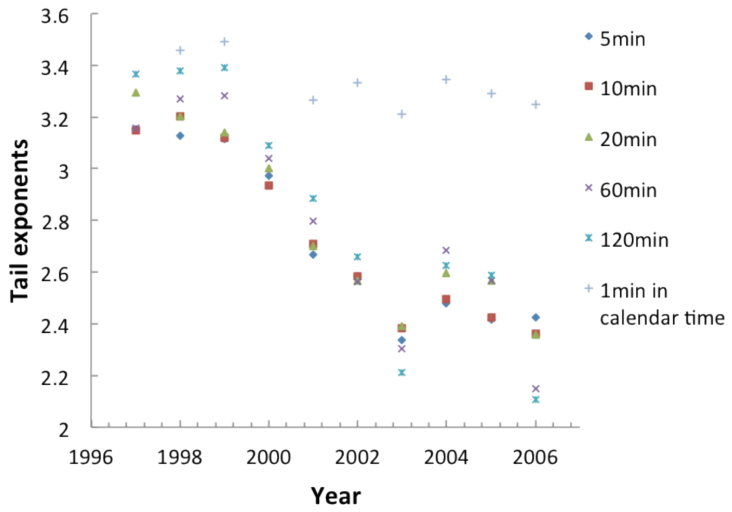


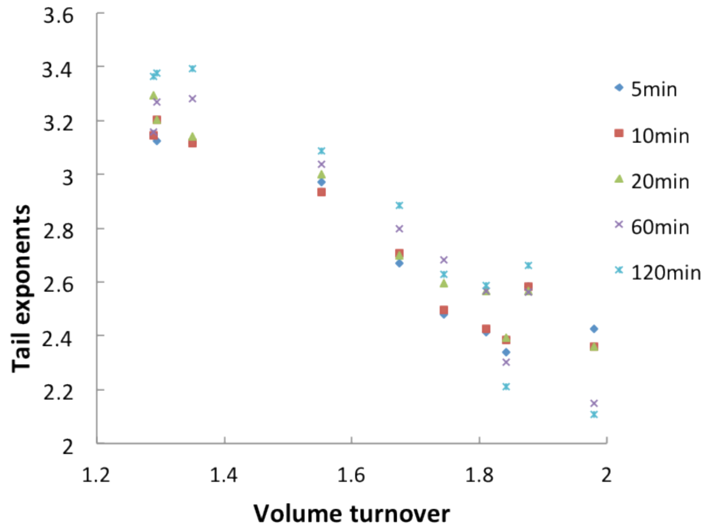
Figure 3.3: Tail exponents for distributions on trade-by-trade returns. For each year, there is no evident variation in tail exponents with respect to different τ.

It is interesting to note that tail exponents from trade-by-trade data generally decreases from 1997 to 2006. In figure 3.4a, we can see that the tail exponent increases slowly from 3.4 to 2.2 through the ten-year period of trade-by-trade data. The trend is even clearer if we plot the tail exponent against the average volume turnover for each year as sown in figure 3.4b. The reason behind this trend is unclear.

It could also be noticed that daily return distribution has a fatter tail than day



(a)



(b)

Figure 3.4: Tail exponents for distributions on trade-by-trade returns v.s. different a) years, b)volume turnover. As the volume turnover increases, the tail of return CDF becomes fatter.

return distributions. This is caused by overnight returns. As shown in figure 3.5, tail exponents from daily returns are consistently larger than the ones of day returns, yet smaller than night returns. This is consistent with previous finding in [65].

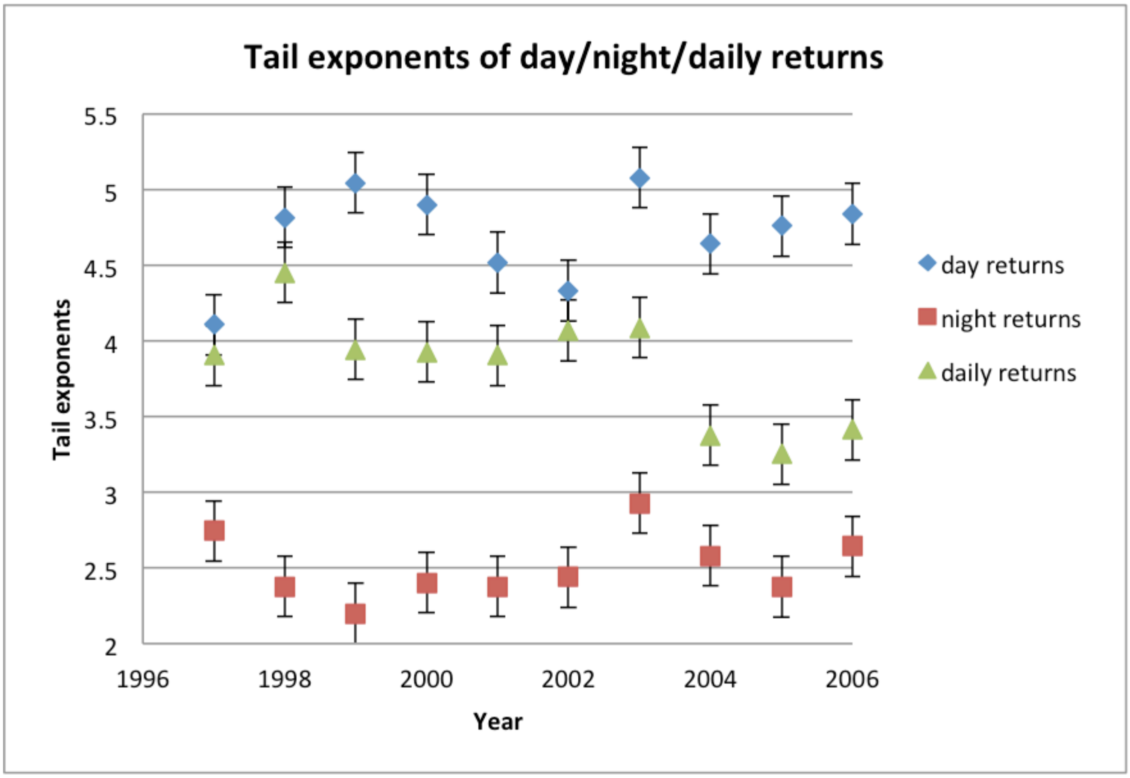


Figure 3.5: Tail exponents for day/night/daily returns. Day returns have thinner tail with larger ξ_r values, while night returns have thicker tail. Daily returns are the sums of day and night returns, and have the tails in the middle of the two.

3.3 Review of empirical data

Night returns occur outside trading hours, when many of the major announcements are made to avoid market impact. Hence overnight return is mostly driven by exogenous factors like macroeconomic conditions, corporate earnings, and government interventions. On the contrary, it is possible that day return is more likely to be driven endogenously. Hence it can be said that day return is closer at describing the endogenous market behaviors with less external influences.

After a thorough examination of return statistics at different frequencies, the trend in tail exponents from high frequency to daily interval is found to be consistent throughout the ten year period studied, and some possible explanation could account for this trend:

1. Aggregate normality. Since the variance of return statistics is finite, decreasing observation frequency is likely to make the distribution closer to normal distribution. On the other hand, the convergence to normal distribution happens rather slowly at intraday level.

2. Market microstructure could be influencing the returns at high frequency level. To understand this one has to study the details of order book statistics and dynamics, which is beyond the scope of this study. A detailed study on order book dynamics can be found in [44].

3. Variation in intraday trading activity. As trading activity has been know to be peaked at the beginning and the end of each trading day [66, 67], this pattern contributes considerably to the autocorrelation of return statistics as pointed out by [66]. Hence it is likely that the return distributions are different at different frequencies.

For the last point of intraday pattern in intraday trading activity, an analysis is demonstrated in the end of next chapter to show how this factor influence the tail behavior of return distributions, through an agent-based model approach.

Chapter 4

Agent-Based Model of Opinion Convergence by Technical Traders

For any agent-based model to be realistic and relevant, it is crucial to have realistic agent behaviors the model parameters. Unlike physics in which controlled experiments can be carried out, in economics and finance it is almost impossible. There has been efforts trying to carry out artificial market to examine how market participants behave [55, 68]. While they provide very useful insight into constructing ABM's, the real market has vastly different types of participants, extremely large agent population, and may behave completely when large sum of money is at stake. Hence much of the work has been done to decipher the underlying market behaviors from the observable quantities like returns or trading volume statistics. Another approach is market surveys. Surveys on fund managers on their trading behaviors can directly extract the important ingredients in understanding market dynamics and fluctuations.

In this chapter, I first start by examining the market ecology and study the different categories of players and their contribution to trading volumes in market place. This is followed by reviewing some of the agent behaviors from both empirical and theoretical evidence, based on which an agent-based model is constructed. Next, parameterization methods and the simulation results are discussed, with additional extrapolation of the result from daily to intraday frequency. Much of the work in this chapter and the next chapter is concisely described in [69]. More details and extensions are presented here.

4.1 Market ecology

4.1.1 Fundamentalists and technical traders

Financial market consists of different players, but most of them can be broadly categorized into two types according to their strategies - fundamentalists and technical traders (as known as chartists). Fundamentalists are usually pension funds, mutual funds and large asset management funds. They are mostly long term investors whose investment positions do not undergo big changes for months or even years. Their investment strategy is driven by fundamental values of a company. Technical traders on the other hand, are usually short term investor who do not pay much attention to fundamentals of a company. They focus on the fluctuations of prices and try to exploit any arbitrage patterns that emerge. Technical traders tend to focus on short term returns and do not hold stocks for a long period.

In the past half a decade, there is a new type of agents emerging in market place, and they are called high frequency trading firms (HFT). As the name HFT suggests, they are the agents who buy and sell stocks at very high frequencies, usually at every few minutes, seconds or less. Studies have shown that the market trading volume contributed by HFT has grown from an negligible amount to more than 50% of the total market transaction today. In this thesis HFT is ignored in the analysis since the period of financial data used is before HFT become significant. But this does not mean the study is not relevant for today's interest, as HFT is mostly used at intraday level and may not affect daily price dynamics.

For the set of 309 stocks studied, majority of the stocks are held by institutional owners as show in figure 2.1. This majority holding represents the fact that most of the outstanding shares are not being traded frequently as institutional players are mostly fundamentalists who hold their stocks for a long term investment.

Although most of the outstanding stocks are held by fundamentalists, it does not reflect the trading activity contributed by them. In order to get a better picture of market ecology, we use the trading volume data to estimate the relative trading activity of both types of agents.

4.1.2 Volume turnover

Volume turnover is retrieved from COMPUSTAT database. In COMPUSTAT, the total number of outstanding shares and total number of traded shares are recorded for each month. For stock i in one particular year, the trading volume turnover is:

$$G^i = \sum_{i=1}^{12} G^i_j / O^i_j, \tag{4.1}$$

where G^i_j is the total number of shares traded in month j of stock i, and O^i_j is the number of outstanding shares in month j for the same stock i. In each year, the average of volume turnovers for all stocks is calculated by

$$V = \sum_{i=1}^{309} G^i / 309. \tag{4.2}$$

In essence, V represents the average activity of the market; it shows how many times of the total outstanding shares are changing hands each year. As shown in figure 4.1, The average volume turnover has increased from 1997 to 2006. From year 1997 to 2002, there is a increase in volume turnover. This is possibly due to the implementation of electronic trading platform that makes trading more accessible with lower costs. The drop in year 2003 happens after the collapse of the dot-com bubble, and trading activity picks up again in 2005. The average volume turnover throughout the ten-year period is 1.64.

4.1.3 Trading volume contribution

Since fundamentalists and technical traders have very distinct holding periods, assuming the average trading velocity (how many times one player changes his/her holdings in one year) by fundamentalists is V_f and the one of technical traders is

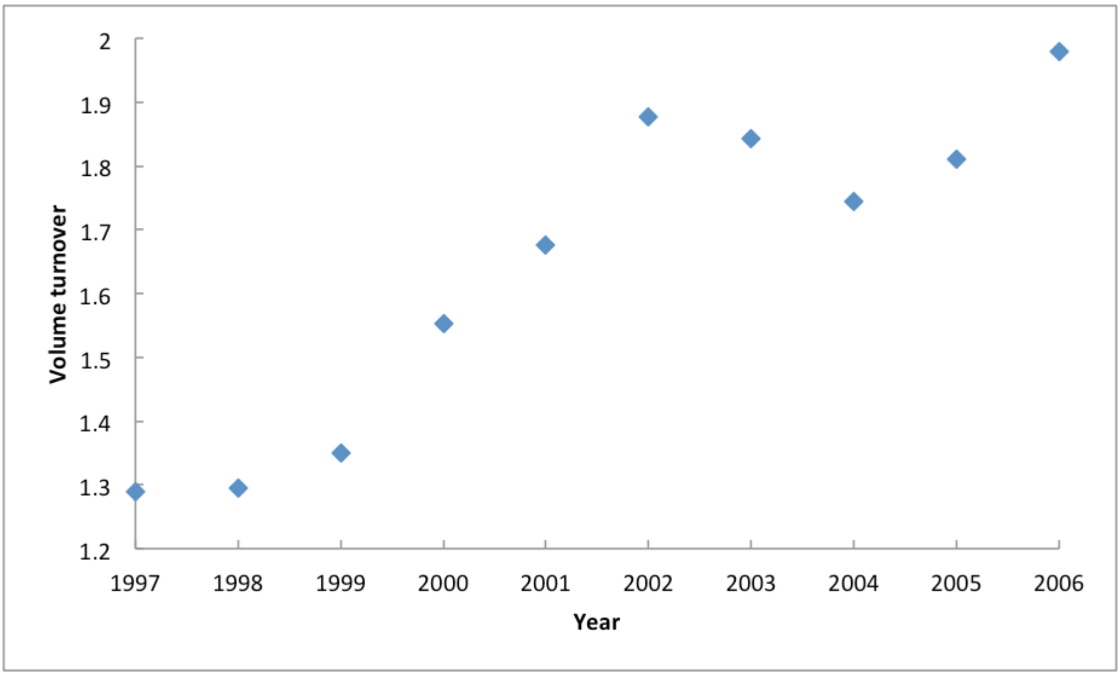


Figure 4.1: Average volume turnover of the 309 stocks from year 1997 to 2006. Trading activity has been increasing through the 10 year period.

V_c, we have the relation:

$$V = 0.83V_f + 0.17V_c \tag{4.3}$$

This means the total trading volume turnover equals to the sum of the contributions by both fundamentalists and technical traders weighted by their holdings of outstanding shares. The ratio between the total number of shares traded by fundamentalists and technical traders is simply $\frac{0.83V_f}{0.17V_c}$. Because fundamentalists have a typical investment horizon longer than one year, their trading velocity is smaller than 1, i.e. $V_f < 1$. Assuming fundamentalists have a investment horizon of 2.5 years, i.e. $V_f = 0.4$, we have $0.17V_c = 1.308$. This means technical traders contribute four times the trading volume by fundamentalists. This result is the opposite of stock holding ratios, and can be seen from figure 4.2.

Although fundamentalists hold a majority of the stocks, they trade infrequently. In contrast, technical traders contribute most of the trading activities [70] by trading their minority holdings much more frequently than fundamentalists. Hence at daily level, technical traders contribute much more to the dynamics of daily price fluctuations than fundamentalists do. This is supported by market survey results on different financial markets [71–73]

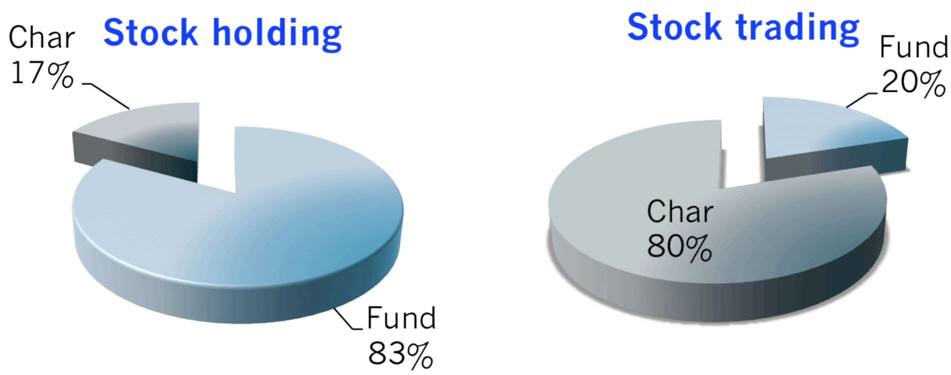


Figure 4.2: Percentage of share holding and trading volume contributed by fundamentalists and technical traders. While the stock holding chart shows that fundamentalists hold majority of the outstanding stocks, the stock trading chart shows the opposite – chartists contribute most of the trading volumes due to their short holding periods.

4.2 Empirical and theoretical agent behaviors

Although the practical usefulness of technical trading has been debated for a long time, its adoption by market participants is undeniable. As demonstrated in the previous section, technical traders play a dominant role in daily trading activities, therefore a dominant contribution to daily price fluctuations. A recent survey on global fund managers [71] shows that technical analysis is widely adopted by fund managers with significant importance at investment horizons up to a year or longer. Similar survey results [72–76] resonate with this fact as well, even though they are carried out in different markets.

Hence in our analysis, we focus mainly on the behaviors of technical traders. As fundamentalists are influenced by exogenous factors like news events and company's earning reports which are unpredictable, they behave in accordance with the random exogenous factors. This contrasts with the conventional wisdom of treating fundamentalists as the main drivers of market fluctuations and technical traders as noise that is purely random. We believe the dominant trading patterns actually come from technical traders yet fundamentalists can be effectively treated as noise that play a less significant role. In simple terms, while conventional wisdom grounds the studies on figure 4.2A, we believe figure 4.2B is a more accurate representation of market.

From this point onwards, two things need to be noted in the rest of the thesis.

- **'Agents' refer to 'technical traders' unless otherwise specified.**
- **All arguments are based on the assumption that high frequency trading is ignored as argued in section 4.1.1.**

Now we can we can gather the relevant agent behaviors before constructing a realistic model.

I **Random trading decisions made by agents on daily basis.** The technical traders use different trading strategies, hence their decisions to buy, sell or hold a position appear to be random. [77] has discovered that on very high frequencies where time interval is less than 60 minutes, no persistency (measured in terms of Hurst exponents [78]) is observed in trading activity. Market survey [71] also shows that fund managers put very little emphasis on intraday trading. Hence it can be assumed that trading decisions are made at daily frequencies for each agent. We estimate the probability p for each agent to have a daily trade from empirical trading volumes, and its calculation is demonstrated in the next section.

II **Opinion convergence** It has been documented that [24] large imbalance in the total number of buy and sell trades usually occurs when most trades carry the same sign. This suggests that big price fluctuations in market are associated with traders executing the same buy/sell action, implying convergence of opinions. In other words, large price fluctuations are not simply due to overall high trading activities, but the fact that most traders have the same opinions on their trading decisions - opinion convergence.

III **Centralized interaction mechanism - price** The handful of most successful agent-based models suffer from similar weaknesses. In the heterogeneous agent-based model [29], fat tail in return distribution disappears as the total number of traders increases beyond a few thousands [46, 79] or interaction structure deviates from random networks [80], unless interaction strength increases with increasing agent population. In the percolation model [42], the average number of interacting neighbors - connectivity probability - needs to be at the critical value to exhibit a power-law tail. These imply there is a centralized interaction mechanism that every agent focuses on, so that this interaction grows naturally with agent population, and is unaffected by interaction structure. For technical traders, an important parameter in their strategies is past price movement. Consequently, prices reflect a main factor through which agents interact. Therefore we propose that price is the centralized interaction mechanism for agents, as it plays an irreplaceable role in most technical strategies.

IV **Same signal of price, diverse strategies, opinion convergence due to price changes** This is the unique mechanism that distinguishes our model from other models. It specifies the collective behavior of technical traders. [81] found that agents learn from each other and tend to adopt the strategy that gives the most payoff. Given the price patterns at any point in time, a few most profitable technical strategies dominate the market because every technical trader wants to maximize his/her profit by using the most profitable strategy copied from each other (the most profitable strategy would become less profitable when most agents adopt it, and a new profitable strategy emerges from the new price trends; soon agents will flock to the new profitable strategy until it is no longer profitable. This is similar to the regime switching phenomenon in various agent-based models). On the other hand, the individual strategies used by different technical traders differ in their parameterizations of the buy/sell time, amount of risk tolerated, or portfolio composition as indicated by [70]. Hence when the input signal – previous price change r_{t-1} – is small, every agent acts independently. When the input signal is large, the agents act more in concert, irrespective of their difference in trading strategies. During the spreading panic of market crashes, for instance, most agents sell their stocks (and market makers are likely to make losses in such circumstances). This supports the empirical finding that large price swings occur when the preponderance of trades have the same buy/sell decision indicated in the findings by [24].

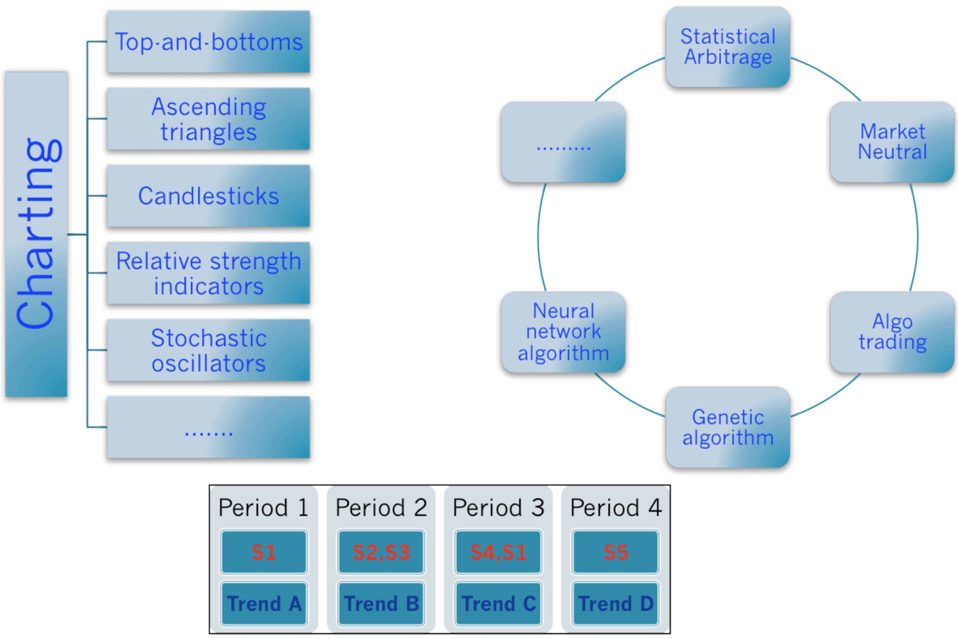


Figure 4.3: An illustration on evolution of technical strategies. Technical strategies have been evolving with time, from the traditional charting methods to modern algorithmic strategies.

Among the four points mentioned above, it is worth spending more explanation on the last one as it plays a critical role in our model. Figure 4.3 shows the basic idea of strategy evolution. There are many different ways to carry out technical trading, and as the trading paradigm shifted to computerized clearing system, even more trading strategies emerged in the recent years. In period 1, for example, strategy S1 is the most profitable given the statistical properties of price fluctuations at that time. More and more traders would switch to S1 and exploit the statistical arbitrage opportunity. But stock market is a zero-sum game and it does not afford to let everyone to win. Hence the act of massively using S1 removes the desired statistical property of price for S1 to be profitable, and new statistical regularity emerged in the next period. The new statistical regularity in period 2 enables new strategies S2/S3 to be highly profitable, until most agents switch to S2/S3 till they are no longer profitable. This cyclic pattern implies that at every point in time, there is only one dominant strategy in the market place - except for the time when market switches from period 1 to period 2.

With all of these being said, the mechanism mentioned above mimics an evolutionary system with ‘profit’ selection. The unique thing about this ‘profit’ selection is that the newly selected strategies always destroys the profitabilities of themselves due to the ‘zero-sum’ nature of financial market. This mechanism is actually the same as ‘strategy switching’ in the heterogenous ABM in [29]. [82] has explicitly modeled the financial system from a strategy-evolution aspect and indeed found that the market has been constantly switching between dominant strategies. In a broad sense, this corroborates with the ‘adaptive market hypothesis’ in [83].

Although one single strategy is likely to dominate the market at any given time, there are different ways of implementing it. Hence under tranquil market conditions where large price fluctuations are rare, the agents are still behaving as if they are using different strategies. This could be due to the fact that they have different risk appetites and leverage ratios, bought the stock at different times, or any other idiosyncrasies they may have. When the market fluctuates drastically, the differences in implementations of the same strategy would be effectively wiped out, and everyone would get the same trading signal from their respective strategies. This effect is similar to the basic ways of recognizing a pattern. When the information people receive is vague, there would be many different opinions on what an object is; when the information is very clear, everyone thinks the same.

4.3 Model construction

In order to investigate the dominant mechanism, we simplify the marketplace and assume every trader is of the same size of unity. This means when they trade, each agent either buys or sells 1 unit of stock. As fundamentalists act on exogenous events, their contribution is considered as noise. Argument *I* leads us to fix the trading time step at daily level. From arguments *III* and *IV*, we propose that the opinions of technical traders converge/diverge when previous price change is large/small. When the price change is extremely large, all of them would be acting unanimously by taking the same trading decision; everyone will act on his/her own when the previous price change is minimal. The exact model has the following three steps:

- Step 1. There are n_0 agents, each of equal size 1. Each day a trading decision $\psi_i(t)$ is made by each agent i,

$$\psi_i(t) \equiv \begin{cases} 1 & \text{with probability } p \Rightarrow \text{buy;} \\ -1 & \text{with probability } p \Rightarrow \text{sell;} \\ 0 & \text{with probability } 1-2p \Rightarrow \text{hold.} \end{cases}$$

  The value of p can be estimated empirically later.

- Step 2 Return r_t is proportional to the aggregate demand d_t, i.e., the difference in the number of agents to buy and sell,

$$r_t \equiv kd_t = k \sum_{i=1}^{n_0} \psi_i(t). \tag{4.4}$$

Trading volume is equal to the total number of trades in this case because every agent has the same trading size of 1. Hence daily trading volume N_t is defined as

$$N_t \equiv \sum_{i=1}^{n_0} |\psi_i(t)| \tag{4.5}$$

and k is the sensitivity of price change with respect to d_t. We set k to 1 because a choice for k does not affect the statistical properties. Note that, according to Step 1, maximum (minimum) d_t means that all agents are in collective mode when they behave the same, all of them either buy or sell stocks.

- Step 3. Based on arguments III and IV, at day $t+1$ each agent's opinion is randomly distributed into each of the c_{t+1} opinion groups where

$$c_{t+1} \equiv n_0/|r_t|, \tag{4.6}$$

in which all agents comprising the same opinion group execute the same action (buy, sell, or hold) with the same probability break down as in Step 1. Since $1 \leq c_{t+1} \leq n_0$, Eq. (4.6) implies (a) when the previous return is maximum, $|r_t|_{max} = n_0$ (everyone buys/sells), there is only one trading opinion among the agents; (b) when the previous return is minimum $|r_t|_{min} \leq 1$, there are n_0 opinions and every agent acts independently of one another. When market noise is considered, $c_{t+1} \sim \mathcal{N}(n_0/|r_t|, \sigma_c^2)$, where $\mathcal{N}$ denotes the normal (Gaussian) distribution and $\sigma_c^2 = b \cdot n_0/|r_t|$ quantifies market noise due to external news events.

A more general form for equation 4.6 is

$$c_{t+1} = (n_0/|r_t|)^\omega \tag{4.7}$$

Equation 4.6 is the special case of $\omega = 1$ in equation 4.7. We will discuss the rationale for choosing $\omega = 1$ in the following section.

4.4 Model justification and parameter determination

4.4.1 Rationale for each step

Despite the lengthy arguments in section 4.2, the model is rather simple, especially the first two steps.

Step 1 is the same as many other ABM including the percolation model [42] as well as the threshold model [43]. But instead of using an arbitrary time step, we define it as one day for our argument I. Since the trading frequency is fixed at daily level, the probability of trading p can be empirically estimated. This avoids the arbitrariness of the choice of parameter.

Step 2 is also a common step among ABMs. The linear price impact in equation 4.4 due to the number of trade imbalance is justified by previous literature findings. Figure 4.4 shows the empirical result of price impact function from [84]. It has found out that for return interval τ = 195min, the price impact by trade imbalance is mostly linear. Trade volume imbalance is the same as aggregate demand d_t in step 2 since each trade are of equal size in our model. It is worth nothing that this linear relationship is not valid at short return intervals as indicated by [22, 84], i.e. small values of τ. In the case of τ is one day, the linear relationship can be justified.

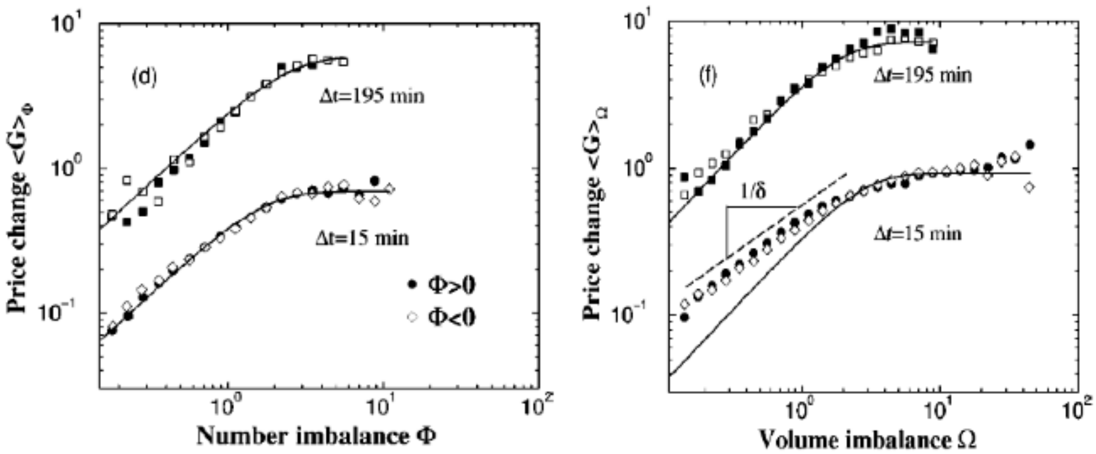


Figure 4.4: Price impact function of left) number of trade imbalances right) volume imbalances. At large time scales *tanh* function is a good fit to the curves. We could see that most part of the curve is linear except for extremely large imbalances. Hence for simplicity of the model it is safe to assume a linear relationship at daily levels. Diagram is obtained from [84].

Step 3 is the most important step in the model, as it mathematically describes the proposed agent behaviors. It specifies that when price change is maximum, i.e. $|r_t| = n_0$, we have $c_{t+1} = 1$, and all the technical strategies would converge to the same decision on buy/sell/hold. Hence every agent would be taking the same action. This means under extremely large price signal, the different parameterizations of different technical strategies are indistinguishable from each other. When price change is minimal, i.e. $|r_t| \leq 1$, we have $c_{t+1} = n_0$, and every agent would have a different opinion based on their trading strategies, hence acts independently. σ^2_{t+1} characterizes the noise term in c_{t+1}, and the noise is assumed to be from either exogenous news events, or endogenously due to randomness in cluster formations; its value is assumed to be on the same scale as the mean value of clusters $n_0/|r_t|$, hence taken to be $b \cdot n_0/|r_t|$.

4.4.2 Determination of parameters

Estimation of p

From section 4.1.1, we have already seen that technical traders contribute to most of the trading volumes, and their trading velocity is related to the average trading turnover by equation 4.3. Hence the average trading velocity of technical traders is given by

$$V_c = \frac{V - 0.83V_f}{1 - 0.83} \tag{4.8}$$

As there are approximately 250 trading days in a year, the daily trading probability p can be estimated as

$$p = \frac{V_c}{250 \cdot 2} \tag{4.9}$$

Because fundamentalists have investment horizon longer than one year, $V_f < 1$.

By choosing V_f to be 0.2, 0.4, 06, 0.8, the value of p is calculated accordingly. The corresponding values of V_f and V_c are listed in table 4.1. From the table we can see that even if the value of V_f varies by four times from 0.2 to 0.8, p changes only about 30% from 0.0174 to 0.0115. This is largely due to the fact that fundamentalists contribute only a small fraction of total trading volume.

Table 4.1: Trading velocities of fundamentalists and technical traders. The probability of having a trade each day by any technical trader is p, and it is calculated from the trading frequency V_f of fundamentalists.

V_f	0.2	0.4	0.6	0.8
p	0.0174	0.0154	0.0134	0.0115

Choice of parameter ω

In equation 4.7, which is a the more general case of equation 4.6, we define the total number of opinion clusters at time $t+1$ as c_{t+1}, and it is directly determined by the previous price changes as $c_{t+1} = (n_0/|r_t|)^\omega$. Since there is at least one opinion among the agents, and that happens when every agent does the same buy/sell action, we have the boundary condition that

$$c_{t+1,min} = 1 \ when \ |r_t|_{max} = n_0,$$

which fulfills $c_{t+1} = (n_0/|r_t|)^\omega$ for any value of ω.

Since there are n_0 agents in the market, there are at most n_0 opinion groups in the market. According to equation 4.7, this occurs when

$$
\begin{aligned}
c_{t+1,max} &= n_0 \\
(n_0/|r_t|_{min})^{\omega} &= n_0 \\
|r_t|_{min} &= n_0^{(\omega-1)/\omega}.
\end{aligned}
$$

This means whenever $|r_t| = |r_t|_{min} = n_0^{(\omega-1)/\omega}$, the market would have the maximum of n_0 opinions and every agent acts independently.

When $\omega = 1.0$, we have $|r_t|_{min} = 1$, and it means when there is exactly one more buy/sell trade than sell/buy trade in the market in the previous day, the market will have totally diverse opinions.

When $\omega = 0.5$, we have $|r_t|_{min} = n_0^{-1}$. We have $c_{t+1} = \sqrt{n_0} \ll c_{t+1,max} = n_0$ (for large n_0) when $|r_t| = 1$, meaning the market has some convergence in opinions when the previous return is extremely small. Under such circumstances, the market tend to converge opinions much easier. Even when return r_t is at noise level of $\sqrt{n_0}$, the number of opinion cluster is very small $c_{t+1} = n_0^{1/4}$.

When $\omega = 2.0$, we have $|r_t|_{min} = \sqrt{n_0}$. This means the market would have totally diverse opinions whenever $|r_t|$ is smaller than $\sqrt{n_0}$, which is the size of noise level. In such cases the opinions of agents tend to stay diverse, and convergence would almost never occur.

Hence the choice of $\omega = 1.0$ is more realistic in mimicking the opinion convergence/divergence mechanism as convergence does happen but not too often.

Other parameters

For parameter n_0 which represents the total number of agents, we typically choose a large number and see if the result would be affected. We also simulate the model for different values of b to see what is the impact of different noise level on simulation outcomes. k is set to 1 as argued in step 2.

4.5 Simulation results and analysis

In order to check the validity of our modeling results, we use the aggregate distribution for the 309 stock as benchmark. Here we compare both the distribution of returns and total trading volumes. Since the model assumes time step of one day, daily return and trading volume data are used from the 309 stocks. For returns, day returns are used instead of daily returns. Day returns are simply daily returns without the component of overnight returns. Since we focus on the endogenous effect by technical traders, by taking away overnight returns which occur outside trading hours, the exogenous influence of news events is reduced as argued in chapter 3. We believe day return is a better reflection of technical trading interactions compared to daily returns.

Simulation is carried out in $1,000,000$ time steps, which is in the same order as empirical data points of approximately 800,000. When tail exponents are evaluated for return distributions, Hill estimator is applied on top 1% of the tail region to estimate tail exponents of returns since power-law is valid only for that part of the tail. In the case of daily number of trades, the distribution has a longer power-law tail. Hence we use 2.5% of the tail region for Hill estimator.

Figure 4.5 shows a comparison between empirical and simulation results. Here simulation is carried out with the parameter values of $V_f = 0.4, n_0 = 2^{10}, b = 1.0$. It can be easily seen that the simulation result is very close to the empirical distributions. ξ_r from empirical data is 4.08 ± 0.08, and 3.69 ± 0.07 for simulation results. In the case of trading volume distribution, ξ_N from empirical data is 4.02 ± 0.07, and 4.02 ± 0.08 for simulation results.

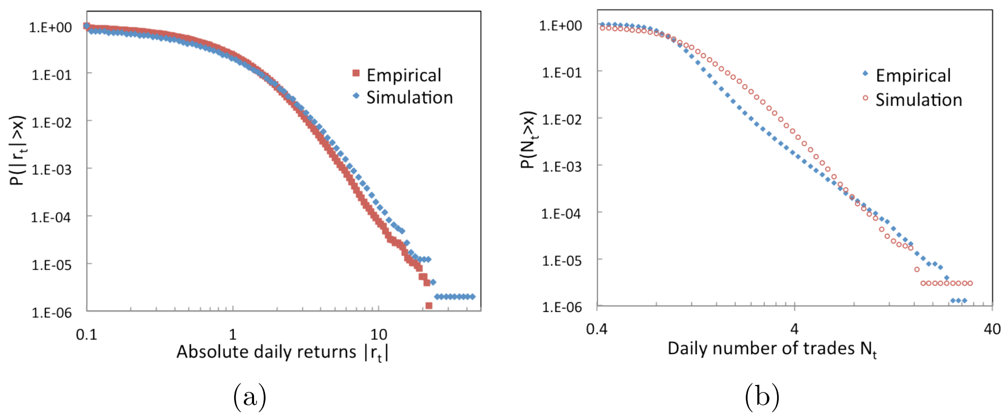


Figure 4.5: Comparison between simulation results and empirical data from S&P500 daily returns. a) Log-log plot of CDF for absolute returns. Simulation shows a power-law tail close to empirical distribution with similar exponents. b) Log-log plot of daily number of trades. Tail behaviors are similar among the two curves with similar exponents.

4.5.1 Different values of V_f

To investigate how the parameter V_f affect simulation result, we change the value of V_f according to table 4.1, and apply the corresponding values of p in the model. Other parameters stay the same: $n_0 = 2^{10}$ and $b = 1.0$. From figure 4.6, we could see that for the four sets of values of p ranging from 0.0174 to 0.0115, the tail of absolute return distribution always follows a power-law, and their shapes are very similar to each other, signifying the robustness of model result against parameter V_f. The tail exponents are little changed as shown in table 4.2.

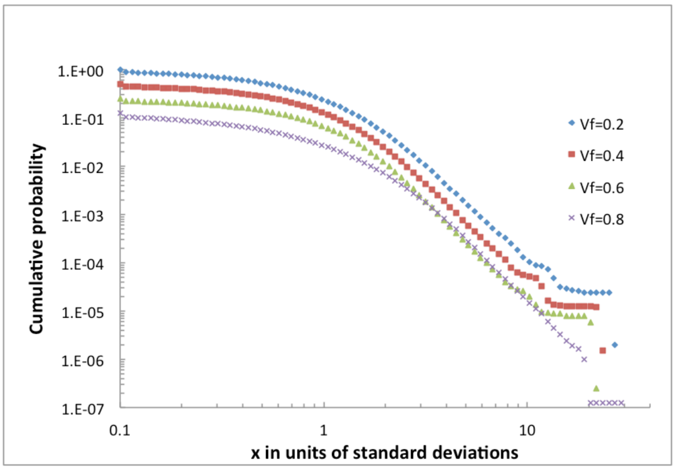


Figure 4.6: Cumulative distribution of returns with different values of V_f in the model. The curves are shifted vertically for viewing. With different V_f values, the CDFs are similar in shape with similar tail exponents.

Table 4.2: Tail exponents from simulation with different values of V_f.

Empirical	Simulation			
4.08 ± 0.07	$V_f = 0.2$ 3.69 ± 0.07	$V_f = 0.4$ 3.69 ± 0.07	$V_f = 0.6$ 3.74 ± 0.07	$V_f = 0.8$ 3.89 ± 0.07

4.5.2 Different values of n_0

In this set of simulations, $V_f = 0.4$, $b = 1.0$, and the total number of agents varies from 2^8 to 2^{14}. As table 4.3 shows, the power-law tail behaviors is almost invariant for different values of n_0. Figure 4.7 shows the distributional plots.

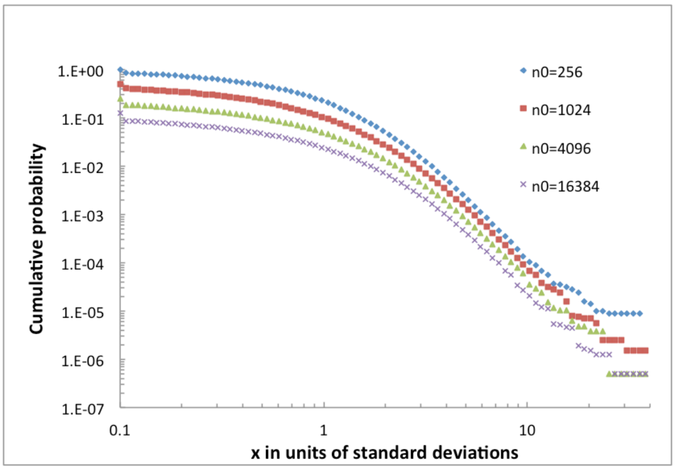


Figure 4.7: Cumulative distribution of returns with different values of n_0 in the model. The curves are shifted vertically for viewing. With different n_0 values, the CDFs are similar in shape with similar tail exponents.

Table 4.3: Tail exponents from simulation with different values of n_0.

Empirical	Simulation			
4.08 ± 0.07	$n_0 = 2^8$ 3.86 ± 0.07	$n_0 = 2^{10}$ 3.70 ± 0.07	$n_0 = 2^{12}$ 3.66 ± 0.07	$n_0 = 2^{14}$ 3.69 ± 0.07

4.5.3 Different agent size distribution

The assumption that every agent are of the same size 1 does not reflect the heterogeneity in the population. It has been reported that trade sizes follow a power-law distribution with exponent around -1.5 [3]. To make the model more realistic, the uniform size distribution of $x_i = 1, \forall i$ is modified to $P(x) \sim x^{-1.5}$ where x is the size of an agent. By following the same simulation steps in the ABM and the same set of parameter settings in figure 4.5, but only changing the size distributions, simulation of the modified model is run again. Figure 4.8 shows the CDF of uniform and power-law agent size distribution side by side. It can be seen that both distribution have power-law tails with exponent around -3. This indicates the robustness of the model that is insensitive to the distributional properties of agent size distributions. Hence, for the simplicity of analysis, the rest of the thesis assumes uniform distribution of agent sizes.

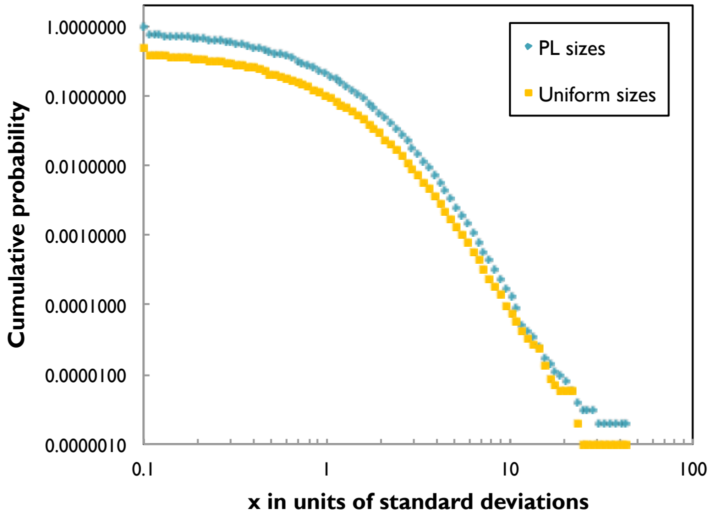


Figure 4.8: Cumulative distribution of returns with uniform and power-law agent size distributions. The distributional property does not affect simulation results as both CDF have very similar shape both at the tail and central regions.

4.5.4 Different values of b

In our model, parameter b characterizes the noise in the number of clusters. To see how this parameter affect our modeling result, we let b to take values from 0 (no noise) to 3 (three times the total number of opinion clusters). From figure 4.9, we can see that when the noise increases, the tail of the distribution becomes fatter. If we interpret noise b as external market information, then it means when there are a lot exogenous impacts on the market, large price swings are more likely to appear, which is in line with casual empirical observations on financial market.

Table 4.4 shows the estimated tail exponents for different values of b. These values are not accurate at large b values as the flattening of the extreme tail region makes the distribution deviates from power-law tail. But it still serves as a indicator on the effect of market noise on price fluctuations.

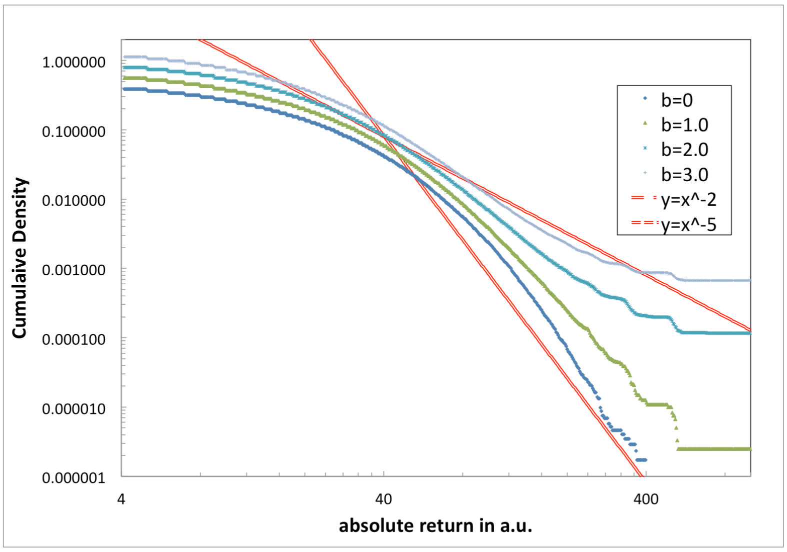


Figure 4.9: Cumulative distribution of returns with different values of b in the model. With increasing noise level characterized by larger b values, the tail of the distributions becomes fatter.

Table 4.4: Tail exponents from simulation with different values of b.

Empirical	Simulation			
4.08 ± 0.07	$b = 0$ 4.12 ± 0.09	$b = 1.0$ 3.61 ± 0.08	$b = 2.0$ 2.91 ± 0.07	$b = 3.0$ 2.26 ± 0.07

Overall, we are able to generate power-law distributions in both returns and number of trades in agreement with empirical data. The fact that we are able to capture the trends in the two highly correlated quantities [85] implies a very plausible mechanism underlying our model. Furthermore, the tail exponent ξ_r is invariant with respect to different values of fundamentalists' investment trading velocity V_f. In additional, our model result is not affected by the increase in total number of agents n_0. This in part solves the well known problem of 'N-dependence' that has been extensively discussed in literature [86, 87].

4.6 Extrapolation to intraday returns

Intraday seasonality in financial market has been well documented in academic literature [16, 88, 89]. In order to understand the stochastic nature of financial time series, in particular volatility clustering, seasonality effect needs to be removed or normalized. Various methods have been proposed to deal with such issue: transformation of physical time scales [90], wavelet filtering [91], transaction time scale [28] and detrended fluctuation analysis [92]. We have seen in section 3.1 that the tail of return distribution is different at different frequencies, and this has been confirmed by earlier studies and was explained in terms of aggregate normality. But The convergence to normal distribution in financial returns is usually very slow, and there is no evidence for why it should happen at intraday scales. Some studies have explained the different tail exponents at different frequencies in terms of the intraday seasonality patterns. In particular, [93] has found in German market that removing the first few minutes - where highest trading activities locates - results in a different tail exponent. Thus seasonality does play a role in the return distributions at intraday level.

Having an ABM allows us to extrapolate our model from daily to intraday returns by taking into account of intraday seasonality trading activities. We divide each trading day into 390 minutes, and study the average trading activity at each minute of the day. Since the trading activity has been increasing from 1997 to 2006, and differs from company to company, we take only Microsoft in 1997 to estimate the intraday trading activity patterns. Figure 4.10 shows the result. There is a clear pattern of inverted 'J' shape in trading patterns throughout the trading day, with the opening minutes having significantly more trades than noon time, and slightly more trading just before the closing minutes. The decrease in activity during noon time is sometimes called 'lunch time' effect.

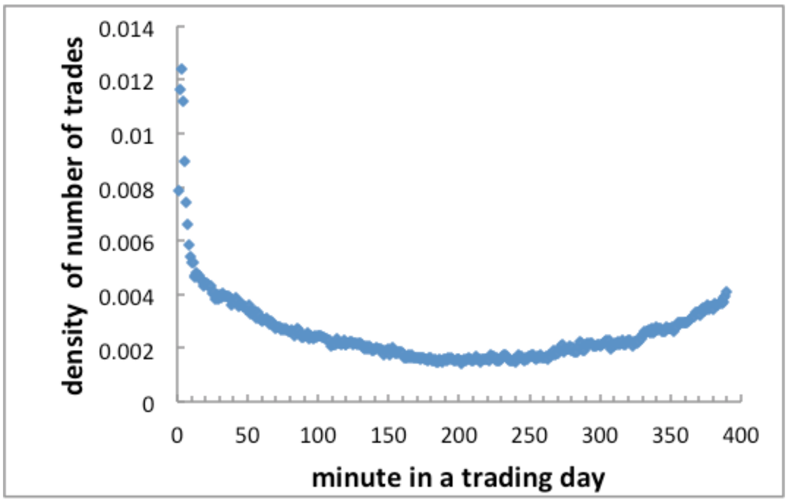
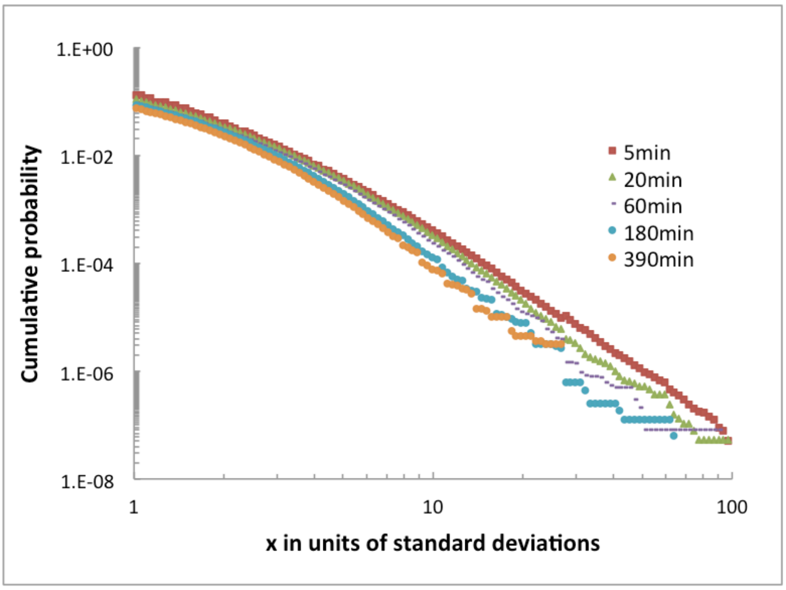
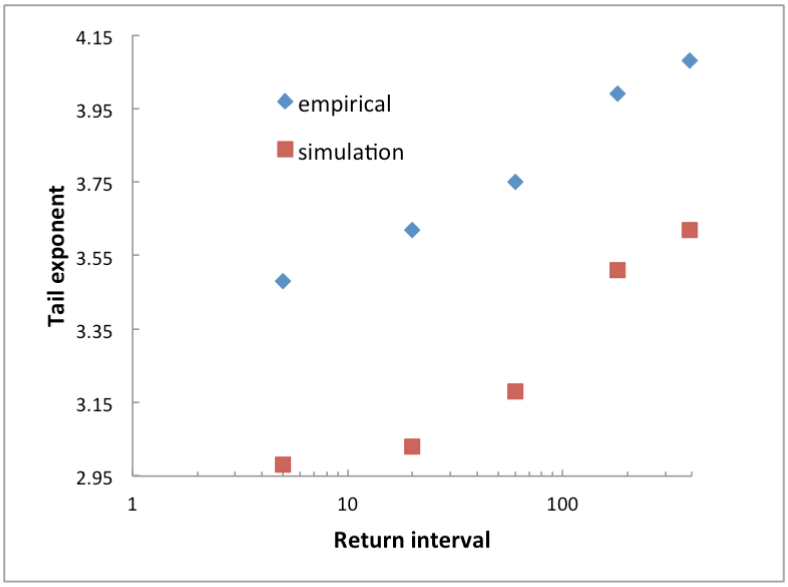


Figure 4.10: Density of number of trade in each minute of a trading day for Microsoft in 1997. Average is calculated on all trading days of the year. Peak of activity is found to be at the beginning of the day and at the end of the day.

In the simulation, we modify the step 2 of the model by breaking day t into 390 intervals, so each interval corresponds to 1 minute in real time. Each agent can choose to trade in any of the 390 intervals, with the probability distribution from figure 4.10. Hence at the end of the day, the total demand at each time interval can be calculated, and high frequency returns are obtained accordingly. In the model simulation, we let n_0 to be a large number of 2^{15}, so that discreteness at high frequencies can be avoided. V_f is still chosen to be 0.4 and $b = 1.0$. The CDF of high frequency returns from the simulation result is shown in figure 4.11a. It is evident that from time scales of 5 minutes to 390 minutes, there is a slight drop in the tail region for CDF. This is in line with our empirical find in figure 3.1. To illustrate this further, tail exponents are calculated from different frequencies and compared with empirical results as shown in figure 4.11b. The increasing trend from empirical data is well captured by the model.



(a)



(b)

Figure 4.11: Comparison of return distributions between simulation and empirical data.(a)Cumulative distribution of price changes at high frequencies from our model simulation results. It is evident that the tail gets thinner when time interval increases. (b)Plot of tail exponents against return time intervals. The increasing trend in empirical data is well captured by the simulation result.

To confirm that this changing tail behavior is not due to aggregate normality or difference in sample size, we check what happens if the intraday pattern is removed from the simulation. Instead of using the trading pattern in figure 4.10, a uniform trading activity is used. This means each agent is likely to execute the trading action in any minute of the day with the same probability. As seen in figure 4.12, the simulation shows no clear indication of changes in tail behavior. Estimation of tail exponents makes this argument stronger as demonstrated in table 4.5, as the tail exponents is almost invariant at different frequencies. Hence we could conclude that intraday trading pattern contributes to the varying tail exponent significantly, if not dominantly.

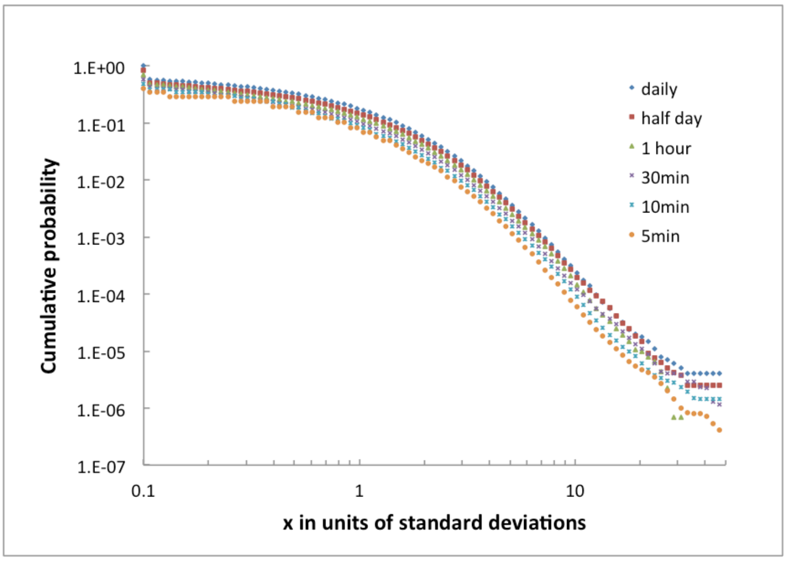


Figure 4.12: CDF plot for simulation at different time scales without intraday pattern of trading activities. The tail behaviors at all frequencies are almost identical.

Table 4.5: Tail exponents at different frequencies for simulation returns without intraday seasonality. The value is roughly the same for different time intervals signifying similar tail behaviors.

τ	390min	195min	60min	30min	10min	5min
ξ_r	3.60 ± 0.08	3.53 ± 0.06	3.56 ± 0.03	3.57 ± 0.02	3.67 ± 0.01	3.88 ± 0.01

4.7 Summary on agent-based model

Starting from the empirical observation that technical traders dominate trading activities, we manage to construct an ABM that is able to capture the cumulative distribution of both returns and daily number of trades. The model has only three steps, and three parameters p, n_0 and b. We can see that the model is able to quantitatively produce realistic distributional properties with different parameters values – the result is robust with a realistic set of values of p estimated empirically, and also not affected by the increasing value in n_0. b characterizes market noise, and the variation in b produces a realistic phenomenon that extreme returns are more likely to occur when market noise is high.

'Universality' of power-law tails has been found in many different financial markets. For an ABM to be successful, 'universality' implies the model needs to be simple, and robust against the reasonable variation in the parameter settings. In other words, the strengths of interactions needs to be insignificant, yet the structure of interaction has to play a dominant role as what has been discovered in phase transitions of physical systems. Our ABM has come close to this success.

While the model is able to capture one of the major stylized facts/universalities, it does not address the issue of long memory and volatility clustering. There are other mechanisms at work in financial markets. Rather than incorporating those mechanisms into our existing ABM, in the next section we demonstrate how to transform the ABM into a mathematical model; and by putting in additional ingredients, how long memory in volatility is quantitatively related to agent behaviors.

Chapter 5

Stochastic Volatility Model

While ABM is able to shed light on the dominant mechanisms behind price fluctuations, mathematical analysis is needed to have a deeper quantitative understanding. There are different ways to do mathematical analysis on ABMs, one of them is the ARCH type analysis. ARCH family has a well-established framework in analytical tools, and many of them can provide a direct comparison to our model. Although the ARCH type models ignore certain non-linearity effects, their simplicity ensures the usability. Sticking to the ARCH formulation also enables us to have a simple analytical understanding of the market dynamics.

In this section, we extend our ABM to a mathematical model and try to demonstrate some of the important stylized facts including fat tail and long memory. Calibration method for the model is presented along with simulation result. At the end of the section, the connection between phase transition in physical system and scaling in financial time series is illustrated.

5.1 Mathematical analysis of ABM

At each time interval in our ABM, return constitutes two major components: a deterministic variance which depends explicitly on past returns associated with equation 4.6, and a stochastic component determined by the random decision of each cluster in step 3 of ABM. Thus we can decouple this two components into σ_t^2 and η_t, such that the daily return is $r_t = \sigma_t \eta_t$, following the formulation of standard ARCH type models. It is worth nothing that return is defined as normalized return after removing the average. Hence it can be treated as the 'residual' term ϵ_t in the ARCH type models, i.e. $r_t = \epsilon_t$ in our case, and we use them interchangeably.

5.1.1 ARCH type model

In the ABM constructed from the previous section, the daily price change is not a deterministic quantity. It has a variance of price change defined as $\sigma_t^2 \equiv \langle (r_t - \langle r_t \rangle)^2 \rangle$, and is related to the total number of opinion groups c_t. The variance in price change r_t is equivalent to variance in total aggregate demand because in step 2 of the ABM the two quantities are proportional (actually they are equivalent in our model construction since the proportionality constant is 1). Since each cluster acts independently, the variance in price change (or total aggregate demand) is the sum of the variances of each opinion cluster's aggregate demand. As each agent randomly chooses a cluster, it can be assumed each cluster has the same size of $|r_{t-1}|$, for there are $c_t = n_0/|r_{t-1}|$ clusters as given by step 3 in the ABM (the variance σ_c^2 due to variation in cluster numbers is ignored for simplicity). Each cluster has probability p of having aggregate demand r_{t-1} (buy decision) or $-r_{t-1}$ (sell decision), and probability $1-2p$ of having aggregate demand of 0 (do nothing or hold). Thus the variance contributed by each cluster is $p \cdot r_{t-1}^2 + p \cdot (-r_{t-1}^2) + (1-2p) \cdot 0 = 2pr_{t-1}^2$. Hence the total conditional variance at time t is

$$\sigma_t^2 \equiv E(r_t^2|r_{t-1}) = c_t \cdot 2pr_{t-1}^2 = 2pn_0|r_{t-1}| \tag{5.1}$$

Knowing the conditional variance of r_t, we can represent r_t as

$$r_t = \sigma_t \eta_t \tag{5.2}$$

Here η_t is a random variable with zero mean and unit variance. We note that the variance σ_t^2 is not constant but time-dependent, and time-dependent variance is commonly found in a variety of empirical outputs where phenomena are ranging from finance to physiology [94]. From this analysis, a possible explanation to ARCH effect in financial data is derived from the behavior of technical traders - larger previous price change r_{t-1} brings traders' opinions closer to each other, resulting in large subsequent price fluctuations.

But there is a caveat here: in ARCH and its related models, η_t is usually assumed to be i.i.d, i.e. η_t at different t are independent and identical distribution. This assumption is not valid in our ABM as we shall see later.

5.1.2 Excess kurtosis

The fact that η_t has a different distribution at different times can be seen from the conditional excess kurtosis. Since each cluster acts independently of each other and has the same size, the total excess kurtosis is $1/c_t$ times the excess kurtosis of each cluster. Thus the total excess kurtosis of the distribution η_t is given by

$$Kurt(r_t|r_{t-1}) \equiv \frac{\langle r_t^4 \rangle}{\langle r_t^2 \rangle^2} = (\frac{1}{2p} - 3)/c_t \tag{5.3}$$

Since c_t can vary between 1 to n_0, and $p \approx 0.015$, the excess kurtosis has a range of $0 < kurt(r_t|r_{t-1}) < 30$ for large n_0. In terms of ABM, this means when the opinions are more converged (agents take similar actions) the kurtosis of η_t is huge, and extremely large returns are more likely; when the opinions are diverged, η_t approaches standard normal distribution.

This is rather complicated as both the variance and kurtosis of the conditional distribution of returns are changing over time, and it is rather difficult to write down explicitly the conditional distribution η_t, thus precise mathematical description becomes extremely difficult. Hence in our study standard normal distribution is used for η_t for the simplicity of mathematical analysis, while simulations are carried out on other distributions to demonstrate the robustness of the model.

The assumption that η_t is i.i.d. standard normal can give rise to unconditional excess kurtosis:

$$\begin{aligned}
Kurt(r) &= \frac{\langle {r_t}^4 \rangle}{\langle {r_t}^2 \rangle^2} = \frac{\langle {r_{t+1}}^4 \rangle}{\langle {r_{t+1}}^2 \rangle^2} \\
&= \frac{\langle \sigma_{t+1}^4 \eta_{t+1}^4 \rangle}{\langle 2pn_0 |r_t| \rangle^2} \\
&= \frac{\langle (2pn_0 |r_t|)^2 \eta_{t+1}^4 \rangle}{\langle 2pn_0 |r_t| \rangle^2} \\
&= \frac{\langle \sigma_t^2 \eta_t^2 \eta_{t+1}^4 \rangle}{\langle |r_t| \rangle^2} \\
&> 3 \langle \frac{\sigma_t}{|r_t|} \rangle^2 = \frac{3\pi}{2} \\
&> 3.
\end{aligned}$$

In the second last step we have used the relation :

$$\langle \frac{|r_t|}{\sigma_t} \rangle = \langle |\eta_t| \rangle = \int_{-\infty}^{\infty} \frac{|x|}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2) = \sqrt{\frac{2}{\pi}} \tag{5.4}$$

In the case that η_t has conditional kurtosis larger than 0 (which is the lower bound as discussed in the previous section), the unconditional excess kurtosis is even larger. Hence our model clearly gives a significant excess kurtosis that is observed in empirical data.

5.2 Heterogenous investment horizons

While traditional economic theories assumes only rational agents, the debate over whether technical traders/technical strategies can survive in the highly competitive market place has been going on over the decades [95]. While technical trading is directly related to arbitrage, the belief in arbitrage-free market has been one of the fundamental assumptions in many of the economic theories including the famous Black-Scholes model [96]. But on the other hand, the assumption of arbitrage-free market is more helpful in deriving analytical solutions than capturing realistic market dynamics.

The best way to tell the extend of usage of technical trading is possibly through surveys on market participants, as it provides a direct and explicit account of how the agents behave. The literature on such surveys has been carried out since the 90s on both foreign exchange market and stock market [71–73], and all of the evidence show that technical trading has been widely adopted by traders across the world. In this study we focus on the survey result on fund managers who invest in stock market, for the empirical analysis of our work is from stock market.

[71] has done an extensive survey from 692 fund managers in five different countries, and the result has shown that up to a forecasting horizon of weeks, technical analysis is the most significant factors in their evaluations, yet fundamental analysis becomes more significant only at longer horizons. Figure 5.1 quoted from [71] gives a graphical representation of this fact.

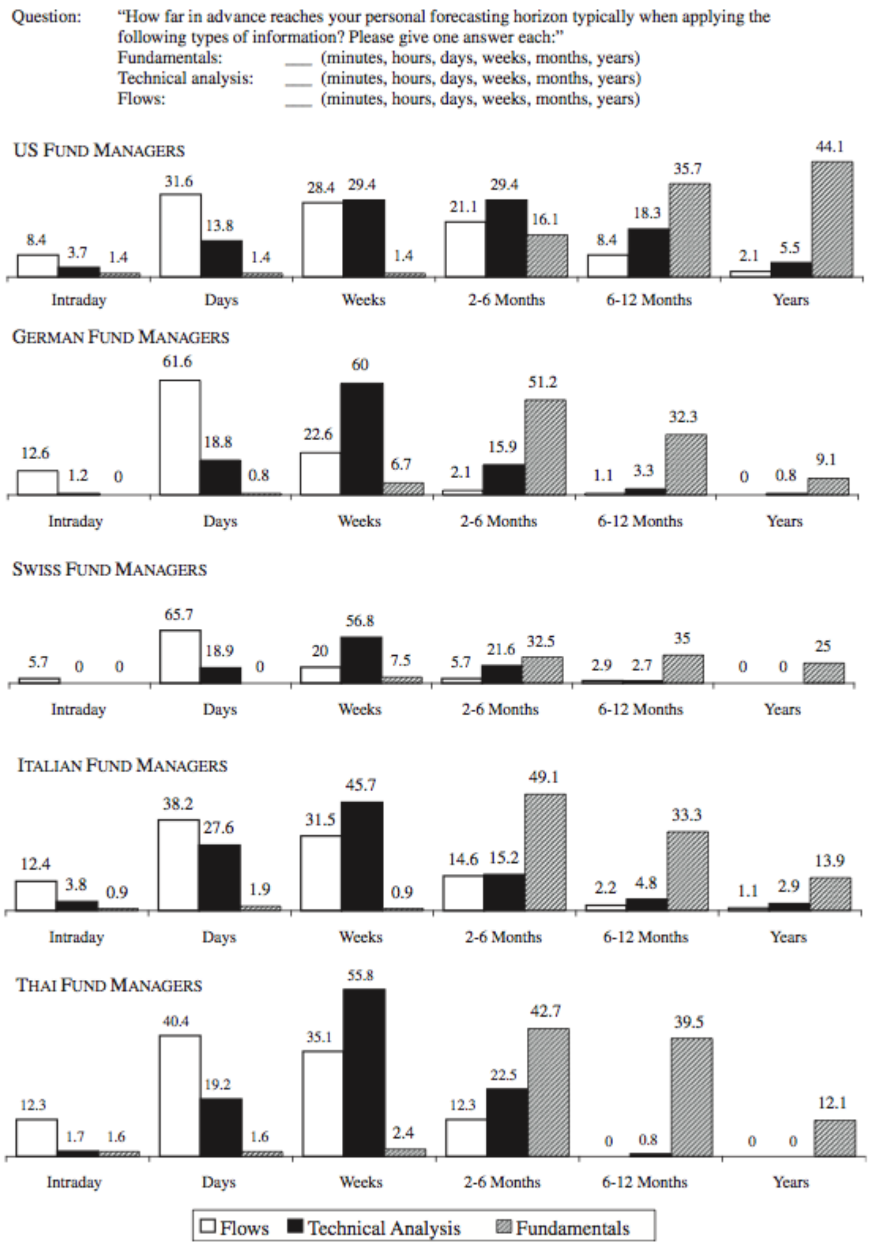


Figure 5.1: The survey on methodologies employed by fund managers in different countries at different time horizons. At intraday level only a small proportion of fund managers rely on any type of analysis – we could say that they dont make decisions at intraday level. At length scale from days to weeks technical analysis dominates the use of fundamental analysis [71].

It is evident from figure 5.1 that at short forecasting horizons up to two months, fund managers predominantly use technical analysis and flows. Flow analysis focuses on volume flows in the market, and can be categorized as technical analysis in a broad sense. For period longer than two months, the importance of fundamental analysis becomes significant and overtakes technical analysis. On the other hand, at intraday level, flow is the dominant analysis and the other two are hardly used.

From this point, we add two additional behaviors of agents on top of the four listed in section 4.2:

V **Technical strategies are applied at different return intervals.** In contrast to the simplistic realization in our ABM where technical traders make their decisions upon the most recent daily returns (4.6), we now assume traders evaluate their strategies based on returns of different past periods. Looking at the survey result of [71] in another way, agents calibrate their technical strategies based on returns of different time horizons, i.e. how a stock performed during the last day, week, month or year. Most technical strategies focus on short-term investment horizons, and fewer on long periods.

VI **Increasing trading activity in volatile market conditions.** Agents tend to trade more after large price movements. Different technical strategies set different thresholds on prices to trigger trading decisions [43], so large price fluctuations are likely to trigger more trades. Hence the probability of having a trade each day p is directly related to past returns. As a large proportion of technical strategies is applied with short investment horizons, price changes from previous days have larger impact on p than from the past year.

5.3 Stochastic volatility model

The two additional assumptions above can be incorporated into equation 5.1 to extend the model.

5.3.1 Heterogenous investment horizons in opinion convergence

For assumption V, we can interpret the differences of applying technical analysis at different forecasting horizons as the differences in the number of technical traders at different investment horizons. Precisely, agents having investment horizon i days are affected by price change in the past i days, and the relative portion of such traders is characterized by α_i. The relative value of α_i can be seen as the relative importance of technical analysis by fund managers in figure 5.1. Let α_i be the percentage of technical traders with investment horizon i, with $1 \leq i \leq M$. M is the maximum investment horizon possible for a technical trader. M is clearly a finite number as technical strategies is not effective on duration longer than a few years when the price pattern is no longer meaningful to be exploited. Thus we have the relation

$$\sum_{i=1}^{M} \alpha_i = 1 \tag{5.5}$$

holds for any combination of α_i. From the survey result in [71], the values of α_i can be realistically estimated.

Since our empirical data is from US stock market, the survey result on US fund managers is used from figure 5.1. As flow analysis is one type technical analysis in a broad definition, we combine the percentages of flow and technical analysis and treat both as the percentage of technical analysis usage at different horizons. The data from intraday level is ignored since all kinds of analysis is little used. For each horizon, the total percentage of technical analysis usage is divided by the length of the horizon (for instance, '2-6' months horizon covers 21 to 120 trading days. Hence the length of the horizon is 100 days) to get an average estimate of α_i. It has to be noted that this is a very inaccurate estimate, but it is the best we could do with the available data at hand. Figure 5.2 shows the plot of estimated α_i at different horizons i. There are only five data points limited by the survey results. It can be seen that α_i decays with i with approximately a power-law function of exponent $d = 1.12$, i.e. $\alpha_i \propto i^{-d}$ with $d = 1.12$. Thus we have

$$\alpha_i = i^{-d}\Omega \tag{5.6}$$

where $\Omega = [\sum_{i=1}^{M} i^{-d}]^{-1}$ is the normalization constant to make sure the total percentage of agents α_i at different investment horizons i sum up to one, i.e. $\sum_{i=1}^{M} \alpha_i = 1$.

Agents with investment horizon i react on the price change in the past i days $|s_t - s_{t-i}|$, hence the opinion convergence due to the technical traders at different i contributes to the total conditional variance σ_{t+1}^2 with a weight α_i. Thus the opinion convergence equation 4.6 in step 3 becomes

$$c_{t+1} \propto \frac{n_0}{\sum_{i=1}^{M} \alpha_i |s_t - s_{t-i}|} \tag{5.7}$$

Since there can be at least one opinion and the maximum return over i days is n_0

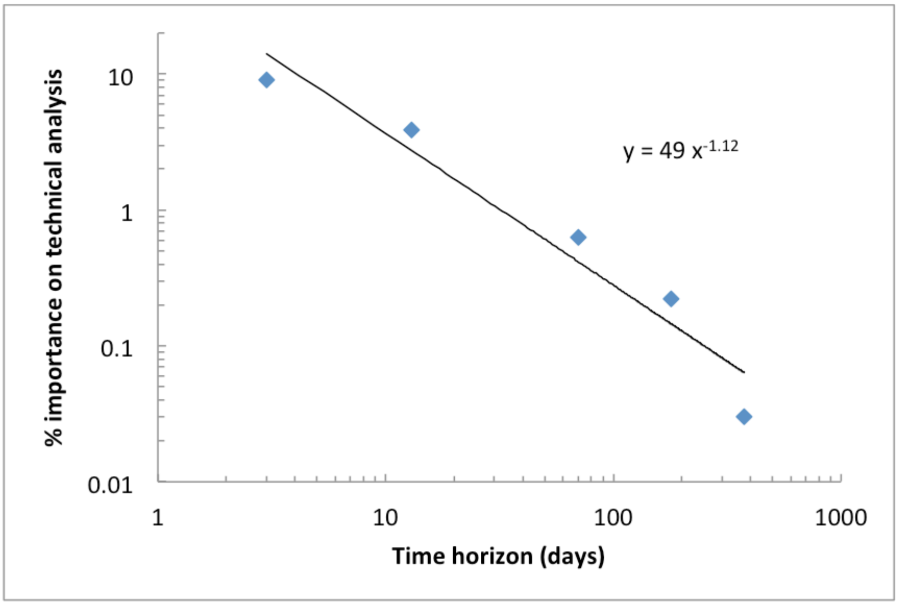


Figure 5.2: Plot of survey result [71] on percent importance placed on technical analysis at different time horizons by U.S. fund managers. The plot shows a power-law function with exponent -1.12. We have combined the percentage values for both flow and technical analysis at each time horizon, as flow analysis in a broad sense is one type of technical analysis. Technical analysis is heavily used at investment horizons of days to weeks and decays to close to zero at 500 days.

(the maximum return cannot be greater than n_0 as this corresponds to the case where all the technical traders have the same long/short position at day $t+1$), this boundary condition requires $c_{t+1} = 1$ when $|s_t - s_{t-i}| = n_0, \forall i$. Thus

$$c_{t+1} = \frac{n_0}{\sum_{i=1}^{M} \alpha_i |s_t - s_{t-i}|} \tag{5.8}$$

Then each cluster has an average of $\sum_{i=1}^{M} \alpha_i |s_t - s_{t-i}|$ number of agents. Thus equation 5.1 becomes

$$\sigma_{t+1}^2 = 2pn_0 \sum_{i=1}^{M} (\alpha_i |s_t - s_{t-i}|) = 2pn_0 \Omega \Sigma_t \tag{5.9}$$

where $\Sigma_t \equiv \sum_{i=1}^{M} i^{-d}|s_t - s_{t-i}|$. Equation 5.9 means the convergence of opinions is not only affected by the previous day's price change, but also the price change of longer past duration in a weighted sum. In the special case of $M = 1$, equation 5.9 gives the exact form of equation 5.1.

Actually equation 5.9 can also be intuitively derived. Since the variance is affected by agents with different investment horizon i, and the size of price change over past i days contributes to this opinion convergence. Conditional variance of price change is a function of $\alpha_i|s_t - s_{t-i}|$, i.e.

$$\sigma_{t+1}^2 = f(\alpha_1|s_t - s_{t-1}|, \alpha_2|s_t - s_{t-2}|, \cdots, \alpha_M|s_t - s_{t-M}|) \tag{5.10}$$

Thus taking the first order (linear) term in the taylor expansion of equation 5.10 we arrive at equation 5.9. This intuitive derivation is a shortcut to derive equation 5.9, but it lacks the detailed behavioral flavor in our ABM.

5.3.2 Impact on trading activity

While we have assumed that the probability of trading each day is constant at p, it is not the case in real market as argued in assumption VI. Several behaviors are associated with this assumption. Technical traders places thresholds on the buy/sell price, so that they can buy a stock once the stock price is desirable, or clear a position to prevent huge losses [43]. Margin call is another factor that forces selling when price movement is too large, so that the margin can be maintained [97]. Order placement is also a factor that contributes the increasing trading activity

[98]. Thus it is more realistic to assume a time-dependent p that is affected by past price changes.

We assume traders put technical threshold around the price i days ago, where i depends on the investment horizons, then the probability of trading p_{t+1} depends on the the past price difference $|s_t - s_{t-i}|$ weighted by α_i as in the previous section. It means larger past price change over the past i days induces more trades for agents with investment horizons i. Thus the probability of trading at day i is

$$p_{t+1} = p_0 + \alpha \sum_{i=1}^{M} (\alpha_i |s_t - s_{t-i}|) = p_0 + \alpha \Omega \Sigma_t \tag{5.11}$$

Here p_0 is the trading probability insensitive to past price changes, and α is a proportionality constant characterizing the size of additional trading activity caused by past volatility Σ_t.

5.3.3 Final stochastic volatility model

With the two additional behaviors analyzed above, equation 5.9 transforms into

$$\begin{aligned} \sigma_{t+1}^2 &= 2p_{t+1} n_0 \Omega \Sigma_t && (5.12) \\ &= 2n_0(p_0 + \alpha \Omega \Sigma_t) \Omega \Sigma_t && (5.13) \\ &= -\frac{n_0 p_0^2}{2\alpha} + 2\alpha n_0 (\Omega \Sigma_t + \frac{p_0}{2\alpha})^2. && (5.14) \end{aligned}$$

The first term is very small compared to the second term because most agents' trading propensities are hugely affected by volatility in market (p_0 very small).

Hence we can ignore the first constant term and focus on the second quadratic term. Therefore the previous equation transforms to

$$\sigma_{t+1} \approx \sqrt{2\alpha n_0}(\Omega\Sigma_t + \frac{p_0}{2\alpha}) = A + B\sum_{i=1}^{M} i^{-d}|s_t - s_{t-i}|. \tag{5.15}$$

where

$$A = p_0\sqrt{\frac{n_0}{2\alpha}} \tag{5.16}$$

$$B = \sqrt{2\alpha n_0}/\sum_{i=1}^{M} i^{-d} \tag{5.17}$$

Equaton 5.15 is the final stochastic volatility model we arrive at, with parameters A and B described by equations 5.16 and 5.17. In this model, every parameter has a realistic interpretation. For equation 4.4 in step 2 of our ABM, the proportionality constant k between aggregate demand and return is set to be 1. In real market the constant k can be merged into A and B without affecting the major statistical properties of the model.

In this model every parameter carries a clear interpretation in real market. It can be seen from equation 5.11 that B/A characterizes the proportion of trades due to past volatility Σ_t; A controls the average size of volatility; M is the longest investment horizon using technical analysis; d characterizes the decays of the use of technical analysis with increasing investment horizon.

5.4 Scaling relations of long memory

To investigate the long memory property of the stochastic volatility model, let us study the property of auto-correlation functions (ACF) in absolute returns and squared returns.

As absolute values are difficult to deal with, we approximate them to

$$|s_t - s_{t-i}| \approx (\sum_{j=0}^{i-1} |r_{t-j}|)/\sqrt{i}$$

Autocorrelation of absolute returns at lag l is given by

$$\begin{aligned}
\rho_l &\sim \langle(|r_t| - \langle|r_t|\rangle)(|r_{t+l}| - \langle|r_{t+l}|\rangle)\rangle \\
&\sim \langle(\sigma_t - \langle\sigma_t\rangle)(\sigma_{t+l} - \langle\sigma_{t+l}\rangle)\rangle \sim \langle(\Sigma_t - \langle\Sigma_t\rangle)(\Sigma_{t+l} - \langle\Sigma_{t+l}\rangle)\rangle \\
&\sim \langle\sum_{i,j=1}^{\infty} i^{-d} j^{-d}(|s_t - s_{t-i}| - \langle|s_t - s_{t-i}|\rangle)(|s_{t+l} - s_{t+l-j}| - \langle|s_{t+l} - s_{t+l-j}|\rangle)\rangle \\
&\sim \sum_{i,j=1}^{\infty} i^{-d-0.5} j^{-d-0.5} \Psi_{ijl},
\end{aligned}$$

with $\Psi_{ijl} = \langle\sum_{m=1}^{i} |r_{t-m}| \sum_{n=1}^{j} |r_{t+l-n}| - \langle\sum_{m=1}^{i} |r_{t-m}|\rangle\langle\sum_{n=1}^{j} |r_{t+l-n}|\rangle\rangle$.

When $j < l$, we have

$$\Psi_{ijl} \approx 0.$$

When $l < j < i + l$, we have

$$\Psi_{ijl} = \Psi_{i,k+l,l} \approx \langle\sum_{m=1}^{k} |r_{t-m}|^2 - \langle|r_{t-m}|\rangle^2\rangle \approx k\sigma_{|r|}^2 \approx (j-l)\sigma_{|r|}^2.$$

When $j > i + l$, we have

$$\Psi_{ijl} = \Psi_{i,i+l+k,l} \approx \langle \sum_{m=1}^{k} |r_{t-m}|^2 - \langle |r_{t-m}| \rangle^2 \rangle \approx i\sigma^2_{|r|}.$$

Therefore

$$\begin{aligned}
\rho_l &\sim \sum_{i,j=1}^{\infty} i^{-d-0.5} j^{-d-0.5} \Psi_{ijl} \\
&= \sum_{i>k}^{\infty} i^{-d-0.5}(k+l)^{-d-0.5} k\sigma^2_{|r|} + \sum_{k>i}^{\infty} i^{-d-0.5}(k+l)^{-d-0.5} i\sigma^2_{|r|}.
\end{aligned}$$

Replacing the summation with continuous integral, we obtain

$$\begin{aligned}
\rho_l &\sim \int_0^\infty \int_k^\infty i^{-d-0.5} k(k+l)^{-d-0.5} didk + \int_0^\infty \int_0^k i^{-d+0.5}(k+l)^{-d-0.5} didk \\
&\sim \frac{1}{d-0.5} \int_0^\infty \frac{d^{1.5-d}}{(k+l)^{d+0.5}} dk + \frac{1}{1.5-d} \int_0^\infty \frac{k^{1.5-d}}{(k+l)^{d+0.5}} dk \\
&\sim \frac{B(1.5-d, 2d-1)}{(d-0.5)(1.5-d)} l^{2-2d} \\
&\propto l^{2-2d}.
\end{aligned}$$

Therefore we obtain $\rho_l \propto l^{2-2d} = l^{-\gamma_1}$, where $\gamma_1 = 2 - 2d$. For autocorrelation in squared returns, we have

$$\begin{aligned}
\rho'_l &\sim \langle (r_t^2 - \langle r_t^2 \rangle)(r_{t+l}^2 - \langle r_{t+l}^2 \rangle) \rangle \\
&\sim \langle (\sigma_t^2 - \langle \sigma_t^2 \rangle)(\sigma_{t+l}^2 - \langle \sigma_{t+l}^2 \rangle) \rangle.
\end{aligned}$$

There are two extreme situations here.

1) If $p_0 \ll 2\alpha' \Sigma_t$, representing the fact that trading tendency due to past price

fluctuations is much higher than background value of p_0, we obtain $\sigma_{t+1} \approx B\Sigma_t$. Assuming that $(s_t - s_{t-i})$ and $(s_{t-i} - s_{t-j})$ are weakly correlated, we obtain $corr(|s_t - s_{t-i}|, |s_t - s_{t-j}|) \approx 0$. Thus

$$\begin{aligned}\sigma_t^2 - \langle \sigma_t^2 \rangle &\sim 2\sum_{i \le j} (ij)^{-d} (|s_t - s_{t-i}||s_t - s_{t-j}| - \langle |s_t - s_{t-i}||s_t - s_{t-j}| \rangle) \\ &\approx 2\sum_{i=1}^{\infty} i^{-2d} ((s_t - s_{t-i})^2 - \langle (s_t - s_{t-j})^2 \rangle).\end{aligned}$$

With similar treatment to ρ_l, we obtain

$$\begin{aligned}\rho'_l \sim \quad & \langle \sum_{i,j=1}^{\infty} i^{-2d} j^{-2d} ((s_t - s_{t-i})^2 - \langle (s_t - s_{t-i}) \rangle^2)((s_{t+l} - s_{t+l-j})^2 - \\ & \langle (s_{t+l} - s_{t+l-j}) \rangle^2) \rangle \\ \sim \quad & l^{4-4d}.\end{aligned}$$

i.e. $\rho'_l \sim \langle (r_t^2 - \langle r_t^2 \rangle)(r_{t+l}^2 - \langle r_{t+l}^2 \rangle) \rangle \sim l^{-\gamma_2}$ where $\gamma_2 = 4d - 4$. For $d = 1.12$, we have $\gamma_2 = 0.48$.

2) In the case that $p_0 \gg 2\alpha'\Sigma_t$, which is approximately the case of our behavioral model, we again have the decay in squared returns as $\rho'_l \sim l^{2-2d}$.

Therefore, the value of γ_2 fluctuates between γ_1 and $2\gamma_1$ depending on the relative size of trades triggered by crossing thresholds placed by technical traders.

Therefore, our stochastic volatility model in equation 5.15 has power-law ACF in both absolute returns and squared returns, i.e.

$$\begin{cases} \rho_l(|r_t|, |r_{t-l}|) \sim l^{-\gamma_1} \\ \rho_l'(r_t^2, r_{t-l}^2) \sim l^{-\gamma_2} \end{cases} \tag{5.18}$$

and the exponents γ_1 and γ_2 are associated with d:

$$\begin{cases} \gamma_1 = 2d - 2; \\ 2d - 2 \leq \gamma_2 \leq 4d - 4; \\ \gamma_1 \leq \gamma_2 \leq 2\gamma_1. \end{cases} \tag{5.19}$$

The scaling relation in equation 5.19 is very similar to the phase transition phenomenon in statistical physics. In statistical mechanics, during a phase transition of a material, certain quantities has a power-law dependence on temperature or magnetism. The exponents in such power-law functions are only related to the structure of the material, and independent of the values of the their interaction parameters. Our stochastic volatility model gives a striking similarity of this phenomenon. It can be seen from our scaling relation in equation 5.19 that the exponents γ_1, γ_2 and d are related in such a way that is independent of parameters A and B in model equation 5.15.

To verify the scaling relations, we test the validity of equation 5.19 on the set of 30 stocks of Dow Jones Industrial Index components. Figure 5.3 shows the plot of the empirical values of γ_1 against γ_2 from each of the 30 stocks, with every point from one stock. The two solid lines represents the relation $\gamma_1 = \gamma_2$ and $2\gamma_1 = \gamma_2$. The scaling relation requires that the points lie in between the two lines. It can be seen that all of the points lie within the two lines except for three of them which still lie on the boarder line of $\gamma_1 = \gamma_2$. It is also observed that the points are spread around $\gamma_1 = 0.4$, which corresponds to $d = 1.2$, and this d value is very close to the empirical value of 1.12 retrieved from the survey result in [71]. Hence we can conclude that the scaling relation predicted by our model can be confirmed by empirical observations.

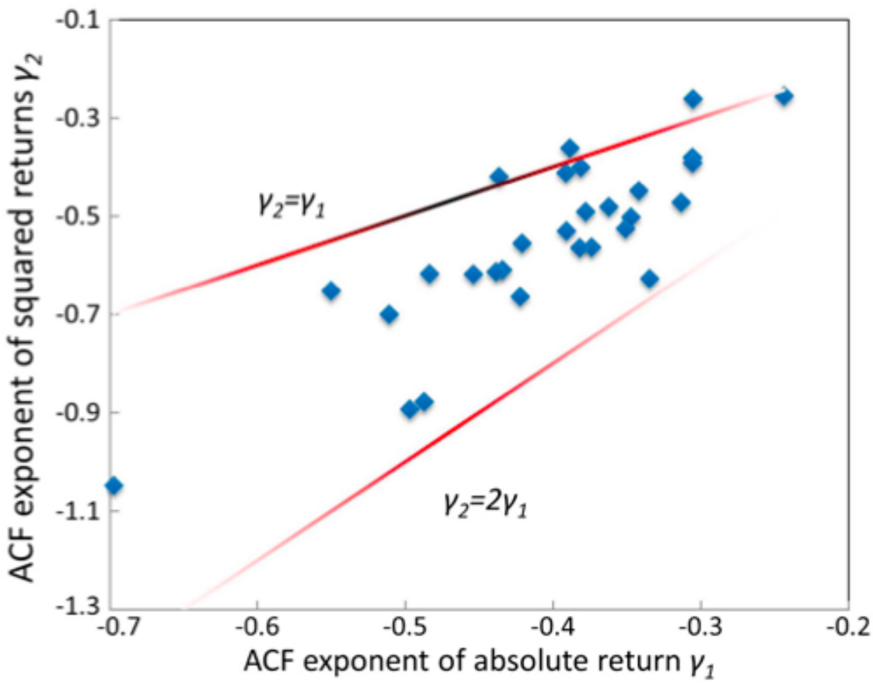


Figure 5.3: Confirmation of the scaling relations in equation 5.19. Dependence of γ_1 against γ_2 for the 30 stocks in Dow Jones Indices components. Daily closing prices from 1971-2010 are used for each stock. The two exponents fulfill $\gamma_1 \leq \gamma_2 \leq \gamma_1$, so the scaling relation predicted by our stochastic volatility model is consistent with these empirical data. The range of lag, $11 \leq l \leq 250$, is used because theoretical power-law decay of the ACF is valid for large l, and for $l > 250$ the ACF values are close to the noise level.

5.5 Monte Carlo simulation

5.5.1 Determination of parameters

The time series model in equation 5.15 has four parameters A, B, d and M. We shall see that most of these parameter carries a realistic meaning and can be retrieved from market.

The parameter d characterizes the decay in the usage of technical analysis with investment horizon, and its value can be found from market survey result. The empirical value is 1.12. In our simulation we use 1.2 which is close to the empirical value.

M is the longest investment horizon that technical strategy is being applied. In the simulation we use $M = 500$ which corresponds to two years. This is in line with the market survey that technical analysis is hardly used beyond one year.

To determine the values of A and B from market data, we look at the ratio of background trading volume p_0 against trading volume due to price impact Σ_t. Since the market landscape has been changing during the last half a century due to computerization in late 90s and the adoption of high-frequency trading in the recent decade, there have been structural changes in trading volumes. Hence we only focus on the 10 year period 1997-2006 during which high frequency trading has not been massively adopted and program trading has already become a norm.

For stock AA, daily trading volume v_i is used from 1997 to 2006 with $i \in [1, 2516]$.

Since trading volume has increased through the ten-year period, we fit the data of v_i v.s. i with a linear function y_i, and form a new detrended time series of $v_i' = v_i/y_i$. From v_i', the average value of the smallest 1% of the data is assumed to be representing p_0, as we know that $p_t \geq p_0$. The average of the whole time series is assumed to be $\langle p_t \rangle$.

From our calculation, $\langle p_t \rangle \approx 3p_0$, which means $\langle p_0 + \alpha\Omega\Sigma_t \rangle \approx 3p_0$. Hence we obtain the relation $p_0 \approx 0.5\langle \alpha\Omega\Sigma_t \rangle$. This indicates trading volume due to price impact is twice of background trading volume p_0. Substituting it into equation 5.15, we get the relation

$$A = \frac{1}{4}\langle B\Sigma_t \rangle = \frac{1}{4}B\sum_{i=1}^{500} i^{-d}\langle |s_t - s_{t-i}| \rangle.$$

Therefore we have

$$\begin{aligned}
\langle \sigma_t \rangle &= A + B\sum_{i=1}^{500} i^{-d}\langle |s_t - s_{t-i}| \rangle \\
&= \frac{5}{4}B\sum_{i=1}^{500} i^{-d}\langle |s_t - s_{t-i}| \rangle \\
&\approx \frac{5}{4}B\sum_{i=1}^{500} i^{0.5-d}\langle |r_t| \rangle \\
&\approx \frac{5}{4}B\sum_{i=1}^{500} i^{0.5-d}\langle \sigma_t \rangle \sqrt{2/\pi} \\
&= \frac{5}{4}\sqrt{2/\pi}\sum_{i=1}^{500} i^{0.5-d}B\langle \sigma_t \rangle. \qquad (5.20)
\end{aligned}$$

Here we have used the identity in equation 5.4. When $d = 1.2$, we have $B \approx 0.05$. The empirical value of $\langle \sigma_t \rangle \approx 0.01$ for S&P500, hence we have $A = \frac{1}{5}\langle \sigma_t \rangle = 0.002$. Analysis on other major stocks gives similar results. It has to be noted that this way of calibrating parameters is very crude, yet it will be seen that it works well in simulation.

It can also be seen that the upper limit in the summation sign of equation 5.20 needs to be finite (in our case it is 500) to have a solution for B, otherwise the summation is not convergent for $d < 1.5$. In the context of the behavioral meaning behind our model, it means technical traders have a finite investment horizon. This is supported again by empirical survey result in [71].

5.5.2 Results

From this point, we have fixed our parameters in equation 5.15 from market data:

$$\begin{cases} A = 0.002; \\ B = 0.05; \\ d = 1.2; \\ M = 500 \end{cases}$$

Daily return r_t is simulated using equation 5.2 as a standard formalism in ARCH type models, i.e.

$$r_t = \sigma_t \eta_t$$

As we have seen earlier that in contrary to standard ARCH models in which η_t is i.i.d., in our model η_t should be time-dependent. But for simplicity we first simulate the model using i.i.d. standard normal for η_t, with the other distributions of η_t tested in the next section.

In figure 5.4 we compare the empirical CDF with simulation result. Similar to chapter 4, the empirical result is the aggregate CDF of all of the 309 stock daily returns. In the simulation 1,000,000 time steps are taken to match the total number of empirical data points. It is evident that the two curves overlap very well, with accuracy up to five standard deviations. Figure 5.5 shows the comparison of ACF between empirical and simulation data. The decays of ACF in both absolute and squared returns show a power-law function against time lag, and this phenomenon is well captured by simulation result. Numerically the simulation result reproduces the empirical decay exponent very well:

$$\begin{cases} \gamma_{1,simulation} = 0.40 & \gamma_{1,empirical} = 0.44; \\ \gamma_{2,simulation} = 0.53 & \gamma_{2,empirical} = 0.70; \end{cases}$$

At the same time, the scaling relation in equation 5.19 is also fulfilled with $d = 1.2$.

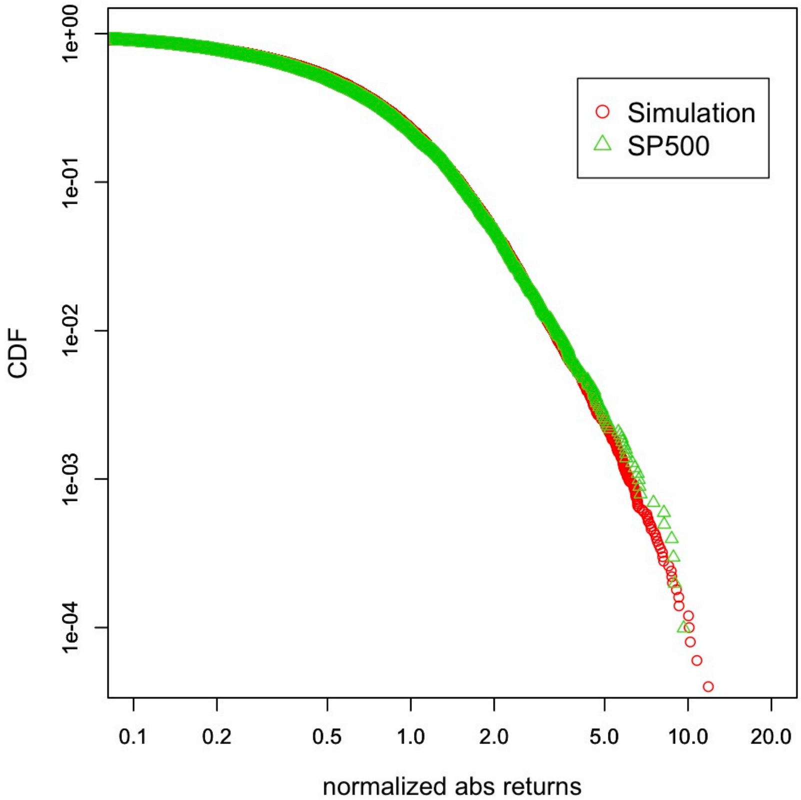


Figure 5.4: Comparison between the CDF of our stochastic volatility model and empirical result of S&P500. The two curves overlap up to 5 standard deviations. Not only the tail part of the distribution is reproduced by the model, the central part is well matched as well.

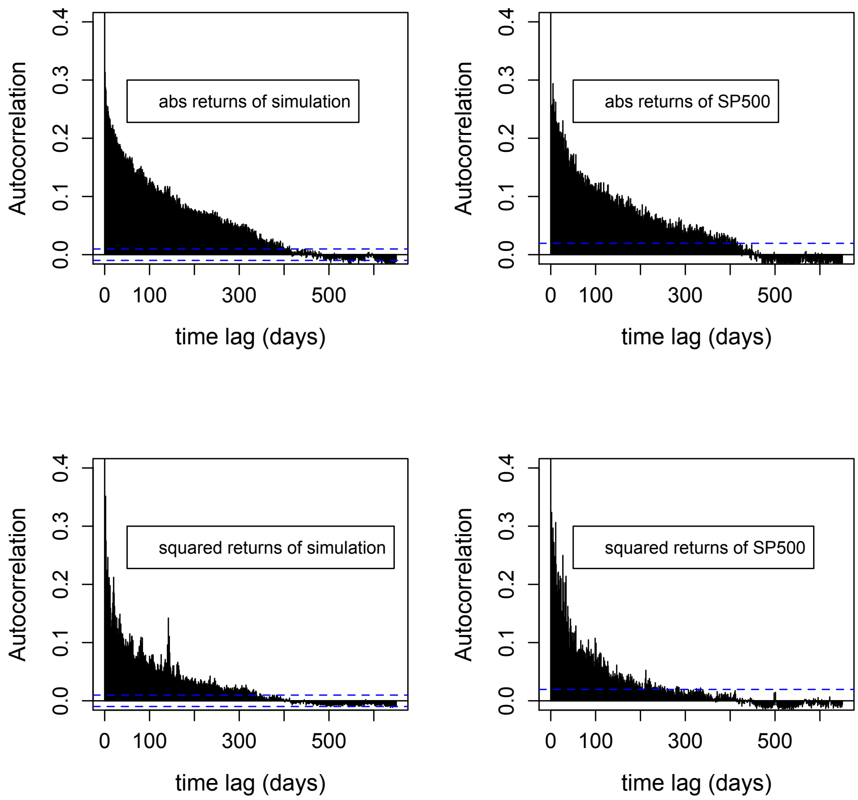
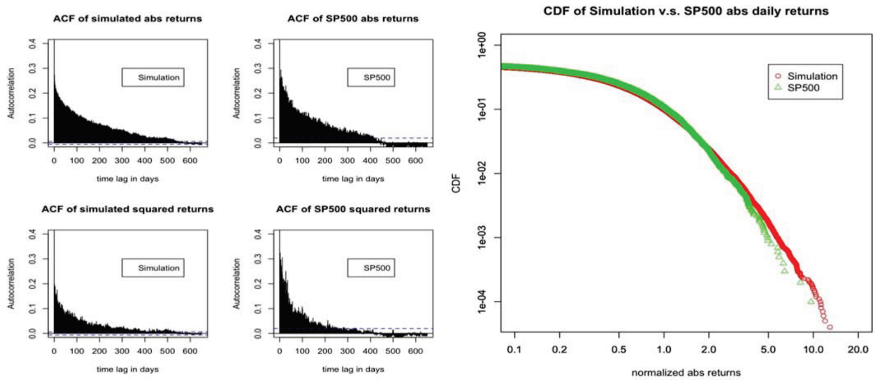


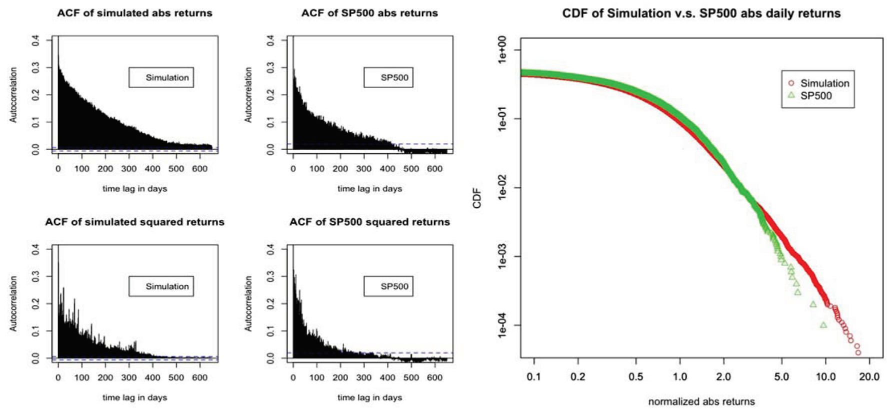
Figure 5.5: Comparison between the ACF of our stochastic volatility model and empirical result of S&P500. The simulation results show similar decay behavior to the S&P500 for both absolute returns and squared returns. The rate of decay is also similar numerically.

5.5.3 Robustness checking

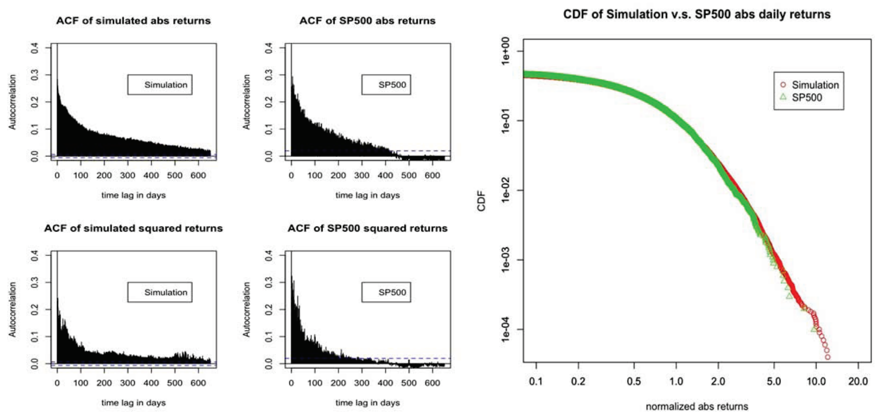
We have discussed that η_t following a i.i.d. standard normal is not a realistic assumption to the agents' behaviors as it does not reflect the excess kurtosis of the conditional distribution. Hence we use Student t distribution of different degrees of freedom to perform simulation. The degree of freedom is taken to be 6, 8, 10 so that the excess kurtosis is $3, 1.5, 1$. It can be seen from Figure 5.6a to 5.6c that the simulation results are still in agreement with empirical finding with different degrees of freedom. The real underlying dynamics has a time varying distribution with excess kurtosis changing between 0 and 30 shown earlier. Since we demonstrate that fixed distributions with different excess kurtosis values give the same outcome, we can safely say the real dynamic process would exhibit the same scaling relations as well.



(a) Degree of freedom 6. Simulation gives $\gamma_1 = 0.38, \gamma_2 = 0.51$.



(b) Degree of freedom is 8. Simulation gives $\gamma_1 = 0.31, \gamma_2 = 0.49$.



(c) Degree of freedom 10. Simulation gives $\gamma_1 = 0.38, \gamma_2 = 0.55$.

Figure 5.6: Plots of simulation results using η_t following t distribution with different degrees of freedom from 6 to 10. Left is ACF plots of absolute and squared returns for simulation v.s. S&P500 index. Right is CDF of absolute returns.

5.6 Long memory and self-similarity

It has been suggested by many works that financial time series is related to critical phenomena similar to phase transitions in natural systems. This is mainly due to the apparent scaling properties in both systems, and their universal scaling exponents. Although critical phenomena in phase transitions have been well understood by renormalization group theory, few works have addressed the origin of the criticality in financial market explicitly. Here from our stochastic volatility model, an analogy between financial market and natural systems can be drawn.

In the Ising model of D dimensions, S_R defines the spin at position R and can only take values of ± 1. At the critical point, the correlation between two spins S_R and S_{R+r} are shown to be

$$G(r) = \langle S_R, S_{R+r} \rangle = K r^{2-D} \tag{5.21}$$

Such a slow decay of correlation with distance r is related to the self-similarity of the Ising model, such that the spin map appears to be similar regardless of the scale of observation.

In the case of financial time series, we can see that the absolute returns at time t and $t+l$ are correlated:

$$\rho_l = \langle |r_t|, |r_{t+l}| \rangle = K' l^{2-2d} \tag{5.22}$$

In financial time series, we observe the scaling properties in time, i.e. the time series looks similar at different time scales: daily, weekly or monthly returns have similar

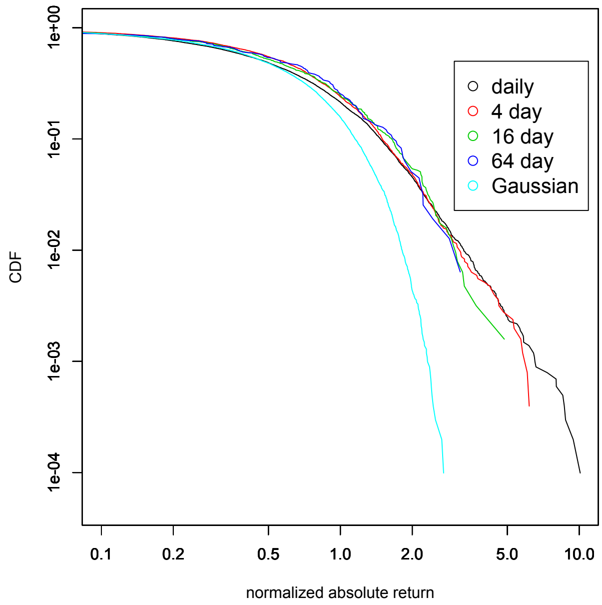
Phase transition	spin S	position R	dim D	$\langle S_R, S_{R+r}\rangle = Kr^{2-D}$
Financial TS	volatility $\lvert r\rvert$	time t	decay d	$\langle \lvert r_t\rvert, \lvert r_{t+l}\rvert\rangle = K'l^{2-2d}$

Table 5.1: Analogy between phase transition and financial time series. The table provides a mapping of variables from phase transition to financial time series.

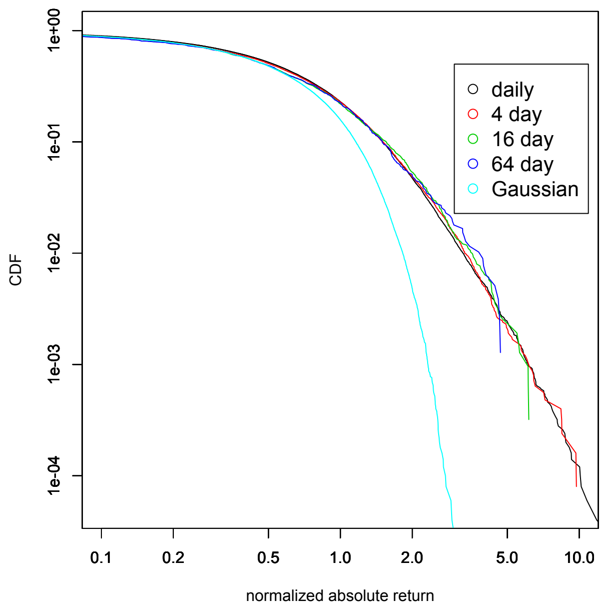
statistical properties. In other words, the self-similarity property of financial time series is closely related to long memory of volatility. If we treat volatility $|r_t|$ as spins S_R, time lag l as length r, and decay exponent $2d$ as dimension D in the Ising model, we have the exact one to one correspondence between the two systems, as shown in table 5.1. The empirical value of $d = 1.12$ corresponds to the physical dimension $D = 2.24$ in phase transition.

In phase transition of the Ising model, self-similarity means the statistical property appears to be invariant with scaling in length. Since length scale corresponds to time scale in financial time series, it means the statistical property of returns is invariant under different time scales - aggregation of high frequency returns. Figure 5.7a shows that the CDF of both S&P500 returns and simulation result from equation 5.15 for different return intervals τ collapse under the same curve. Each curve here is normalized to zero mean and unit variance, and figure 5.8b plots the returns v.s. time for both time series.

This apparent similarity between the Ising model and financial time series has very different underlying mechanism though. In the Ising model, short range interaction results in long range correlations near critical point in the form of power-law decay. The dimension D controls the speed of decay in spatial correlation equation 5.21. In the case of the time series represented by our model equation 5.15, the origin for long range (in time coordinates) correlation is directly related to the long investment horizons of investors. The investors with different investment horizons act as the coupling forces between the returns at different time scales. Such coupling forces decay with distance (again, in time coordinates) as power-law function with exponent d. Hence the origin of such long range correlation is directly related the long range coupling, rather than the short range interaction in the Ising model. In other words, the scalings in phase transition happen as critical phenomena when the system is at criticality; whereas in financial market, it is not so straight forward to see the role of criticality in the system. The investment horizons of agents could be the dominant contributing factor.

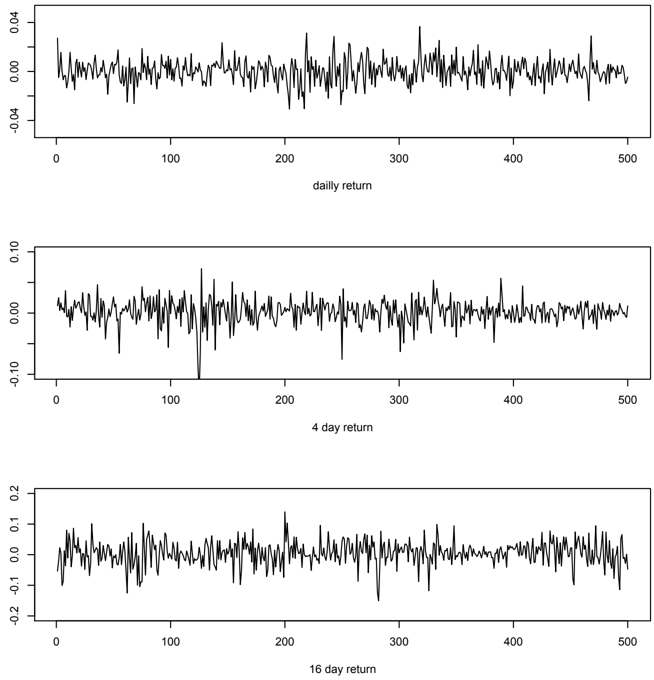


(a) S&P500 returns

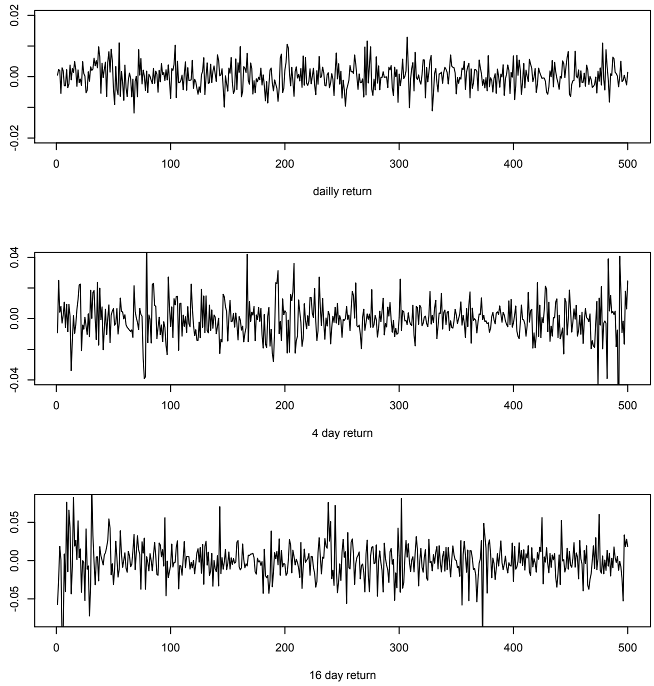


(b) Simulation result

Figure 5.7: CDF of time series of different return intervals for S&P500 and simulation result from equation 5.15. Each time series is normalized to zero mean and unit variance, and they collapse onto the same curve exhibiting similar statistical property. Gaussian distribution is shown side-on-side to demonstrate the fat tail of returns exists at all time scales.



(a) S&P500 returns.



(b) Simulation result

Figure 5.8: Plots of both S&P500 returns and simulation result for different return intervals τ ranging from 1 day to 16 days.

5.7 Summary on the stochastic volatility model

Starting from the underlying mathematical structure of our behavioral model, we manage to derive a stochastic volatility model that is similar to the ARCH construction. Our model construction gives a new approach to calibrate the model parameters: rather than statistical fitting, the parameters can be retrieved from empirical market data with realistic behavioral interpretation behind.

The two stylized facts mentioned in chapter 1 can be directly explained through our model – fat tails in return distributions arise when technical traders converge in their trading strategies and opinions, and long memory in absolute and squared returns are due to the heterogeneous investment horizons of the market participants. Such an explanation is more than a qualitative argument, as mathematical relations can be drawn between empirical behaviors and return statistics as demonstrated.

This is not the first time that a stochastic volatility model is constructed from behaviors, but the mechanism in our model is unique and the accuracy of the model predication has never been achieved before to the best of our knowledge. On the other hand, ARCH family models usually adopt hand-waving arguments of market dynamics, with less behavioral interpretation behind. Although the model shares some resemblance of ARCH family models, there are some key differences, which we will see in the next chapter.

Chapter 6

Relationship to ARCH Models

Since the first introduction of the Autoregressive Conditional Heteroscedasticity (ARCH) model [4] by R. F. Engle in 1982, variations of it have constructed for financial market. The primary aim of such models is to reproduce price dynamics with good accuracy, in particular return distributions, long memory (volatility clustering) and leverage effect.

As our stochastic volatility model has the similar formulation as ARCH type models, it is worth comparing it to the broad range of ARCH models. One fundamental difference between our model and the general ARCH type models is that ours starts from empirically observed behaviors, in contrast to the hand-waving arguments that formulate ARCH models. Putting this aside, we can draw analogies to them on the mathematical level, to see how our model performs and what it means to ARCH type modeling. Since leverage effect is not considered in our model, this chapter does not focus on the models dealing with this particular stylized fact.

In this chapter, residual ϵ_t is equivalent to return r_t in the mathematical analysis, for return in this thesis is defined after removing its mean value as we have argued in the beginning of chapter 5.

6.1 Conditional volatility

Before starting the comparison, let us write down the explicit form of conditional volatility in our model from equation 5.15:

$$\begin{aligned}
\sigma_{t+1}^2 &= (A + B\sum_{i=1}^{M} i^{-d}|s_t - s_{t-i}|)^2 \\
&= (A + B\sum_{i=1}^{M} i^{-d}|\sum_{j=0}^{i-1} r_{t-j}|)^2 \\
&= A^2 + 2A\cdot B\sum_{i=1}^{M} i^{-d}|\sum_{j=0}^{i-1} r_{t-j}| + B^2(\sum_{i=1}^{M} i^{-d}|\sum_{j=0}^{i-1} r_{t-j}|)^2
\end{aligned} \tag{6.1}$$

In ARCH(q) models, the conditional volatility depends on the square values of previous q days' returns, and the dependence is linear as shown in section 2.2.1. While it is the simplest among of all its variations, it has the form which many others can be related to, including GARCH and FIGARCH.

Our stochastic volatility model of equation 6.1 cannot be explicitly related to ARCH(q) models, as the absolute values we are dealing with prevent us from breaking up the individual returns. The other major difference between our model and most other ARCH models is the dependence on the first order past returns.

Some models including Nonlinear GARCH [99] and Threshold GARCH [100] incorporate the first order terms to introduce leverage effect, yet in our model the leverage effect is not reflected since the absolute value of the returns in the equation imposes a symmetrical impact by positive and negative returns.

For the comparisons in the following sections, crude estimations are made in order give close resemblance among the models. Hence it has to be kept in mind that the real dynamics dictated by our model is not as close to others as the comparison appears to be.

6.2 Fitting onto GARCH

Although GARCH models have a volatility memory much shorter than empirical financial data, it is very successful in producing distributional properties. GARCH(1,1) model can be rewritten into a form of ARCH(∞) model as:

$$\begin{aligned}\sigma_t^2 &= \alpha_0 + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \\ &= \alpha_0 \sum_{i=0}^{\infty} \beta^i + \frac{\alpha}{\beta} \sum_{i=1}^{\infty} \beta^i \epsilon_{t-i}^2 \end{aligned} \tag{6.2}$$

In equation 6.2, the coefficient in front of ϵ_{t-i}^2 decays exponentially as $\alpha\beta^{i-1}$, which is much faster than our model which coefficients decay as power-law function. This is the reason behind the fast decaying memory of the GARCH model.

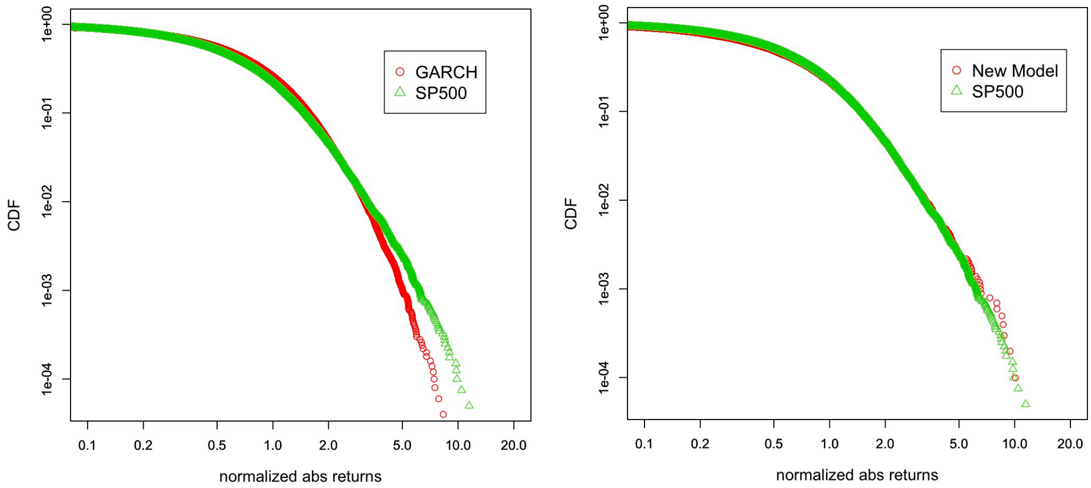
Empirically, fitting a GARCH(1,1) model gives $\alpha+\beta$ smaller but close to 1. Fitting the simulation result of our stochastic volatility model gives the same relationship as well. The fact that $\alpha+\beta \approx 1$ led to the Integrated GARCH model by enforcing $\alpha + \beta = 1$. The problem with IGARCH is that it has a memory much more persistent than empirical data, thus making it less useful in volatility modeling. Being a non-integrated process for our model, the calibration result of $\alpha+\beta$ close to 1 means the apparent integrated process of real financial data may be an artifact of other dynamics.

By fitting the simulation result from our stochastic volatility model equation 5.15 onto a GARCH(1,1) model, the following model is obtained with maximum likelihood estimation(MLE):

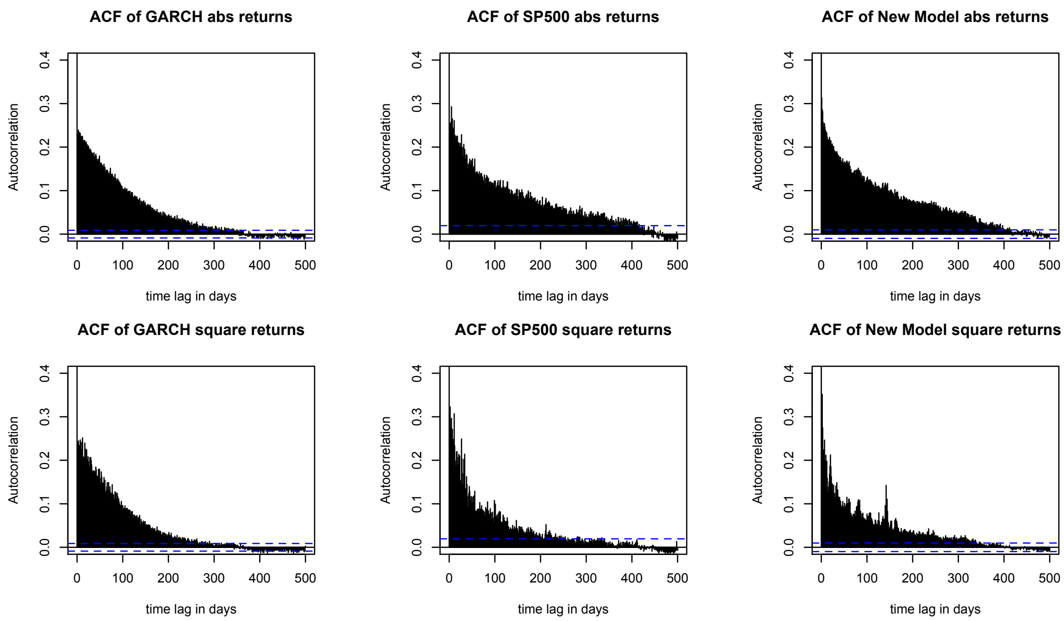
$$\sigma_t^2 = 5.616^{-7} + 0.09375\epsilon_{t-1}^2 + 0.8941\sigma_{t-1}^2 \tag{6.3}$$

Hence the sum $\alpha+\beta = 0.09375+0.8941 = 0.98785 \approx 1$, which is exactly in line with the empirical fact that fitting of GARCH gives a process close to integrated time series. This confirms the accuracy of our model from another aspect: $\alpha + \beta \approx 1$ does not mean the process is close to a integrated one, but may be an artifact of another functional form in our model equation 5.15.

To illustrate the performance of GARCH and our model, Monte Carlo simulations are run for both 5.15 and GARCH(1,1) fitted from empirical data. Without and statistical testing, figure 6.1a shows that our model performs better than GARCH(1,1) at capturing the CDF of absolute returns. It can be seen from 6.1b that GARCH(1,1) is not able to capture the faster decay of ACF in squared returns found in empirical data.



(a) CDF comparison between empirical data (S&P500), GARCH(1,1) and our stochastic volatility model. While GARCH(1,1) is not able to capture the tail distribution closely, our model matches the empirical data closely up to 10σ.



(b) Autocorrelation plots for absolute and squared returns from S&P500, GARCH(1,1) and our model 5.15. While both GARCH(1,1) and our model can capture the decay trend in ACF of absolute returns, GARCH(1,1) is not able to capture the faster decay of ACF in squared returns found in empirical data.

Figure 6.1: Comparison with GARCH(1,1) show that our model outperforms GARCH(1,1) both in turns of distributional and long memory properties.

6.3 Fractionally integrated processes

On a casual look, our stochastic volatility model in equation 5.15 has some similarity to a fractionally integrated process. One particular example of fractionally integrated process was proposed by [101]. In that process, σ_t is written as

$$\sigma_t = \frac{1}{\lambda}(1-(1-L)^{d^{**}})|\epsilon_t| = \frac{1}{\lambda}\sum_{k=1}^{\infty}\frac{d^{**}}{\Gamma(1-d^{**})}\cdot\frac{\Gamma(k-d^{**})}{\Gamma(k+1)}|\epsilon_{t-k}| \tag{6.4}$$

When $k \gg 1$,

$$\frac{1}{\Gamma(1-d^{**})}\cdot\frac{\Gamma(k-d^{**})}{\Gamma(k+1)} \approx \Gamma(1-d^{**})^{-1}k^{d^{**}-1} \tag{6.5}$$

Hence d^{**} is related to our d by $d^{**}-1 = 0.5-d$. In the original paper [101], the value of d^{**} is in the range of $0 < d^{**} < 0.5$, which corresponds to $1 < d < 1.5$, and this covers the empirical value of $d = 1.12$.

Another example of fractionally integrated process is the famous fractionally integrated generalized autoregressive conditional heteroskedasticity (FIGARCH) model [61]. FIGARCH model was developed to work around some weaknesses of the GARCH model. We have seen in chapter 2 that while GARCH model is able to produce the return distribution at the tail end, the autocorrelation property is poorly captured. In particular, the memory of absolute returns in GARCH model decays exponentially, yet empirical data shows power-law decays (or 'hyperbolic' decays as the term commonly used by econometricians). The fact that $\alpha+\beta \approx 1$ leads to the belief returns follow an integrated process, thus IGARCH model was created by fixing the relation $\alpha+\beta = 1$ in GARCH. On the other hand, the time series produced by IGARCH models has a memory much more persistent than empirical data due to the fact that it is an integrated process. Hence a 'mathematical compromise' is made between GARCH and IGARCH models to produce a memory that is neither too short nor too persistent, and the new model was called FIGARCH.

The FIGARCH model representation can be written as:

$$\phi(L)(1-L)^{d^*}\epsilon_t^2 = \omega + [1-\beta(L)]v_t \tag{6.6}$$

where $v_t = \epsilon_t^2 - \sigma_t^2$, L is the lag operator which gives the value of the variable on the previous time interval, i.e. $L\sigma_t^2 = \sigma_{t-1}^2$ and $L^2\sigma_t^2 = \sigma_{t-2}^2$. $\phi(L)$ and $\beta(L)$ are polynomials of L. Equation 6.6 can be transformed into

$$\sigma_t^2 = \omega[1-\beta(L)]^{-1} + \{1-[1-\beta(L)]^{-1}\phi(L)(1-L)^{d^*}\}\epsilon_t^2 \tag{6.7}$$

In the case of FIGARCH(1,d^*,1), we have $\phi(L) = 1$, $\beta(L) = \beta L$, and equation 6.7 becomes

$$\begin{aligned}\sigma_t^2 &= \omega[1-\beta L]^{-1} + \{1-[1-\beta L)]^{-1}(1-L)^{d^*}\}\epsilon_t^2 \\ &\approx \omega^* + \alpha^* \sum_{i=1}^{\infty} i^{d^*-1}\epsilon_{t-i}^2 \end{aligned} \tag{6.8}$$

Thus the conditional variance σ_t^2 has a power-law decay of coefficients with exponent $d^* - 1$.

In our original stochastic volatility model equation 5.14, the conditional variance can be written as

$$\begin{aligned}\sigma_t^2 &\approx \omega^{**} + B^2\Sigma_t^2 \\ &= \omega^{**} + \alpha^{**}(\sum_{i=1}^{M} i^{-d}|s_t - s_{t-i}|)^2\end{aligned}$$

Very roughly, we can say $(\sum_{i=1}^{M} i^{-d}|s_t - s_{t-i}|)^2$ is in the same order as $\sum_{i=1}^{M} i^{2-2d}\epsilon_{t-i}^2$. Hence in our stochastic volatility model,

$$\sigma_t^2 \sim \omega^{**} + \alpha^{**}\sum_{i=1}^{M} i^{2-2d}\epsilon_{t-i}^2 \tag{6.9}$$

Comparing equation 6.8 and 6.9, we can see that the two exponents are related:

$$2 - 2d \approx 1 - d^* \tag{6.10}$$

In practice, FIGARCH model has a value of d^* in the range $0 < d^* < 1$, which corresponds to $1 < d < 1.5$ in our model. This is exactly the range covering our empirical value of $d = 1.12$.

6.4 Heterogeneous ARCH

The heterogeneous ARCH model proposed in [102] is similar to our model in certain ways. It assumes that agents are heterogeneous in a way that they focus on volatility of different temporal resolution. In other words, different traders look at price changes over different horizons. The motivation for assuming heterogeneity comes from the empirical fact that there is a asymmetry in the information flow between fine volatility and coarse volatility. The later predicts the earlier from the lag-correlation results. The authors interpret this empirical finding as the fact that short term traders react to both long term volatility and short term volatility, but long term traders - fundamentalists - react only to long term volatility.

With such an assumption, the conditional volatility is defined as the weighted sum of squared values of past aggregate returns in equation 6.11.

$$\sigma_t^2 = c_0 + \sum_{j=1}^{M} c_j (\sum_{i=1}^{j} r_{t-i})^2 \tag{6.11}$$

Here M refers to the maximum horizon that traders look at, same as the definition in our model. The HARCH model is different from most other ARCH models by using the aggregate returns $\sum_{i=1}^{j} r_{t-i}$ instead of simple returns r_{t-i}. Such a formulation is used because traders focuse on volatility - absolute value of aggregate returns - of different intervals. With the power-law functional form for the coefficients c_j as $c_j = k \cdot j^{\mu}$, the model can give long memory property in simulated time series. i.e.

$$\sigma_t^2 = c_0 + k \sum_{j=1}^{M} j^{\mu} (\sum_{i=1}^{j} r_{t-i})^2 \tag{6.12}$$

$$\tag{6.13}$$

Since the term $(\sum_{i=1}^{j} r_{t-i})^2$ has the same order as $j^2 \sum_{i=1}^{j} r_{t-i}^2$, the comparison with our model yields a relation $\mu + 2 \approx 2 - 2d$, i.e.

$$\mu = -2d \tag{6.14}$$

In [102], the above hyperbolic functional form is not used for calibration. Instead, something called 'market components' are used. 7 values of horizons are chosen in the model separated by powers of p, i.e. $j = 1, 4, 16, 64, 2561024, 4096$, with one unit of time referring to 30 minutes. This corresponds to the interval from 30 minutes to 12 weeks. The argument for choosing 7 components instead of equally spaced 4096 components is that traders of different types usually take typical time horizons, rather than all possible time horizons between 1 and maximum. This assumption resonates with the survey designed in [71], in which 6 typical time horizons are chosen in the survey questions: intraday, days, weeks, 2-6 months, 6-12 months and years.

In the construction of our stochastic volatility model equation 5.15, we have transformed the sparse result of 6 time horizons into a finely spaced 500 time intervals from 1 day to 500 days. It is worth noting that the two approaches share the same numerical origin, and results are expected to be similar as we shall see later.

While HARCH model is able to capture the long memory in volatility, its construction using linear combinations of squared returns is based on 'their good analytical tractability' similar to ARCH family. There is no clear interpretation as how the heterogeneous market components (time interval) are derived from trading behavior. Also, the parameter estimation is carried out using statistical fitting and genetic algorithm, unlike our model which directly uses market data.

6.5 Borlan-Bouchaud

[98] is another very successful model that simulates long memory, and one of its assumption is that agents place stop gain/loss on their positions around the transaction price to limit their risk and lock in the gains. The authors assume the distribution of stop gain/loss price around the transaction price to follow a functional form of

$$P(\Delta) \sim |\Delta| exp(-k\Delta^2) \tag{6.15}$$

where Δ is the distance between the transaction price and stop gain/loss price of the trades at any time. This form seems to be an arbitrary definition by the authors, and it allows easy analytical tractability.

Hence, for trades placed at time $t - i$, the proportion of them being cleared at time t is determined by the the absolute price change between time t and $t - i$. In other words, the additional trades triggered by reaching the stop gain/loss limits contributes to volatility at time t. Under certain approximations, the conditional volatility can be written as

$$\sigma_i^2 = \sigma_0^2[1 + \sum_{l=1}^{\infty} g_l \frac{(x_i - x_{i-l})^2}{\sigma_0^2 l\tau}] \tag{6.16}$$

Here g_l is proportional to the total number of trades with horizon l, τ is the return interval, and g_l is the decay coefficient for the number of trades at different lag l. The authors assume that g_l decays with l in a power-law function, i.e.

$$g_l = g/l^{\mu^*} \tag{6.17}$$

Similar analysis to previous sections yields the relation

$$\sum_{l=1}^{\infty} g l^{-\mu^*} \frac{(x_i - x_{i-l})^2}{\sigma_0^2 l \tau} \sim \sum_{l=1}^{\infty} g l^{1-\mu^*} r_{t-l}^2 \tag{6.18}$$

In a sense the model is very similar the HARCH model. They both depend on the linear combination of the squared aggregate past returns. The additional input by this model is that it is constructed from a more detailed behavioral assumption, though the functional form of stop gain/loss position in equation 6.15 was assumed arbitrarily.

Comparing to our stochastic volatility model, it is equivalent to

$$1 - \mu^* = 2 - 2d \tag{6.19}$$

It is also very interesting that their model gives the decay of auto-correlation in squared returns with power-law exponent $2 - 2\mu^*$, which is exactly what we obtained from $4 - 4d$ by comparing with equation 6.19. For our empirical value of $d = 1.12$, this corresponds to $\mu^* = 1.24$, which lies exactly in the range of $1 < \mu^* < 1.5$ proposed by [98], and is also close to their numerical calibration result of $\mu^* = 1.15$. Apart from this, the model is able to capture time reversal asymmetry of empirical data discovered by [103].

The calibration of the model is called 'soft calibration' by the authors. It means rather than estimating the exact parameter values of the best fitting, approximate parameter values are retrieved to ensure model consistency and reproduction of stylized facts, and it turns out that the fitting is not bad. Although this is one step further than the conventional statistical estimation, in the way that quantitative stylized facts directly infer parameter values, it is still done in a mathematical way, with no connection to empirical values from agent behaviors. The model was also unable to account for the reason behind choosing the memory kennel in the model. It assumes that an infinite kernel length is needed for the model, i.e. the conditional volatility depends on the whole history of the time series information. Yet our model provides an explicit account for the finiteness of the memory kennel, as it depends on the maximum horizon of technical traders, and such an assumption proved to be quite successful in producing good simulation results.

6.6 Quadratic ARCH

The quadratic ARCH(QARCH) model is a general form of ARCH family defined in [104]. The labyrinth of ARCH models are mostly special cases of QARCH, as long as the conditional volatility depends on quadratic forms of past returns.

$$\sigma_t^2 = s^2 + \sum_{\tau=1}^{\infty} L(\tau) r_{t-\tau} + \sum_{\tau,\tau'=1}^{\infty} K_{\tau,\tau'} r_{t-\tau} r_{t-\tau'} \tag{6.20}$$

The linear term is related to leverage effect, sometimes ignored in GARCH and FIGARCH models. The memory kennel $K_{\tau,\tau'}$ comprises the weight of all the combinations of past returns at different times. ARCH, GARCH and FIGARCH only have the diagonal terms in $K_{\tau,\tau'}$, with different decaying trend; HARCH and Borlan-Bouchaud models contains off-diagonal terms that are not negligible.

[31] has provided an in-depth analysis on QARCH models with comparison to other forms of ARCH models. It has found that neither ARCH, GARCH or FIGARCH is good for empirical estimation as they lack non-trivial off-diagonal values in the kennel $K_{\tau,\tau'}$. The reason most models related to diagonal-only kennel performs not too badly is that the off-diagonal entries are significantly smaller than the diagonal ones, yet they are not negligibly small to be ignored altogether. It has also found that none of the existing ARCH related models has the off-diagonal coefficients resembling that of empirical estimation from QARCH model. It means none of the existing models from ARCH family is able to capture the fine volatility feedback property of empirical data in the context of QARCH.

Comparing with QARCH, it is obvious that our model equation 5.15 is not a particular form of QARCH model, for we use absolute values of past aggregate returns that are not possible to be decomposed into individual daily returns. Hence any direct analytical comparison between the two is difficult. One remark that can be made is the decay of diagonal elements in the kennel $K_{\tau,\tau'}$. Estimation of QARCH leads to power-law decay of the diagonal terms, which is in line with models of FIGARCH and HARCH, and in agreement with our model, although strictly speaking our model cannot be decomposed into the kennel representation of QARCH.

It is worth nothing that when applied to stock index, fitting of QARCH leads to significant off-diagonal elements in the kennel at $\tau = 5$ and around $\tau \approx 10$, which corresponds to one week and two weeks trading time. The calibration of QARCH is carried out only for maximum lag of 10, hence any higher τ values are not observed. This finding provides the support for HARCH model [102] using different market components, assuming traders have trading horizons at one day, one week, one months or higher, rather than all possible number of days. This is perhaps a more realistic assumption of agent behaviors, which can be incorporated into our stochastic volatility model.

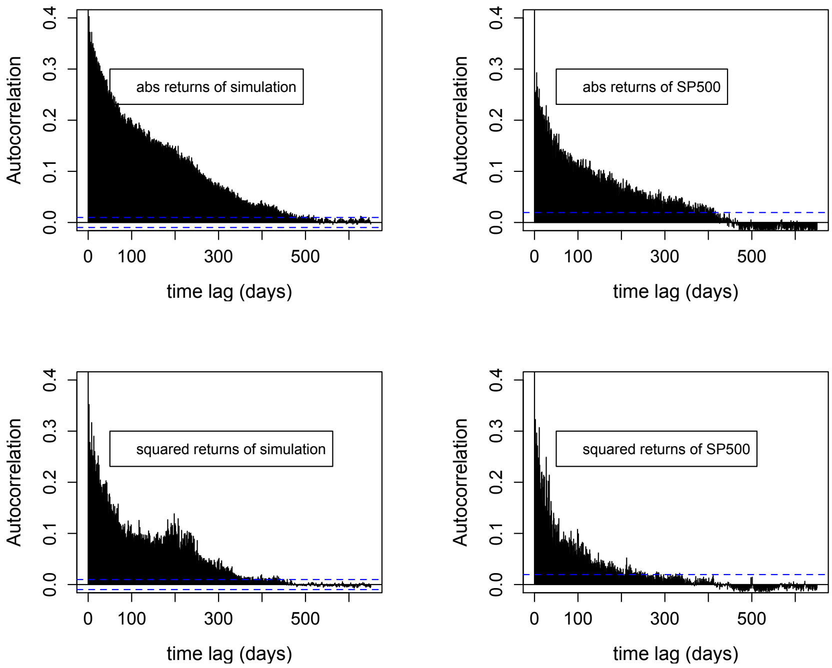
Since the survey questions in [71] are fixed at horizons of days, weeks, 2-6 months, 6-12 months and years, we can choose the market components as $\tau = 5, 20, 80, 200, 500$. Thus our new model can be written as

$$\sigma_t = A + B \sum_{i=5,20,80,200,500} \alpha_i |s_t - s_{t-i}| \tag{6.21}$$

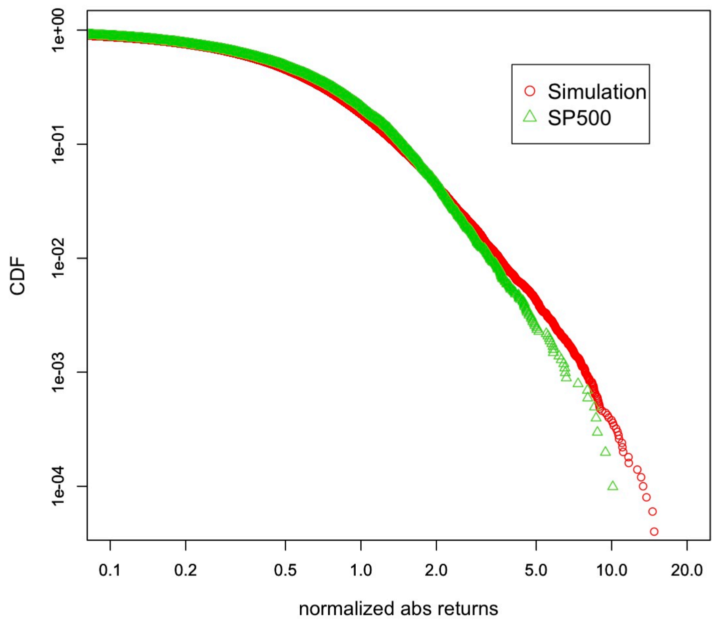
From the same survey result of [71], we can obtain the values of α_i by summing up the percentage of technical analysis and flows. Thus $\alpha_5 = 0.454$, $\alpha_{20} = 0.578$, $\alpha_{80} = 0.505$, $\alpha_{200} = 0.267$, $\alpha_{500} = 0.076$. This is a very rough estimation of parameters, and the values of A and B are estimated similarly to chapter 5 equation 5.20:

$$\langle \sigma_t \rangle = \frac{5}{4}\sqrt{2/\pi} \sum_{i=5,20,80,200,500} i^{0.5} \alpha_i B \langle \sigma_t \rangle \tag{6.22}$$

The corresponding values of A and B are $A = 0.002$ and $B = 0.074$, and the simulation result is shown in figure 6.2. It can be seen that with such a primitive parameterization, our simulation could produce a long memory decay in auto-correlation not far from empirical results, and the CDF is close to empirical up to 5 standard deviations. With more careful selection of lag parameters, the result will definitely be improved. But here we only show an illustration that multiple sparse time lags due to technical trading horizons can be easily incorporated into our model to produce good result, without any statistical estimation.



(a) ACF of 5 components model.



(b) CDF of 5 components model

Figure 6.2: Plots of simulation results of 5 components model with parameters from survey results in [71]. Though the parameters are used from the survey result in a crude way, long memory and fat-tail can be reproduced qualitatively.

6.7 Non-quadratic dependence

It seems that most of the models in ARCH family is written in terms of linear combinations of quadratic forms of past return, and there has been no clear justification on this formulation. Indeed [101] has proposed an alternative formula that is based on the conditional standard deviation instead of its squared value. [105] went even further by defining the value of σ_t^δ with δ being a non-integer:

$$\sigma_t^\delta = \alpha_0 + \sum_{i=1}^{p} \alpha_i(|\epsilon_{t-i}| - \gamma_i \epsilon_{t-i})^\delta + \cdots \tag{6.23}$$

This model is named asymmetric power GARCH by the authors. Here the term $\gamma_i \epsilon_{t-i}$ characterizes the leverage effect. The value of δ is estimated to be 1.43. The rationale behind such non-quadratic models is that auto-correlation is strongest for $|r_t|^d$ with $1 < d < 2$ empirically, thus justifying the choice of d smaller than 2. Possibly due to its difficulty in analytical framework, these two non-quadratic models are not discussed as extensively as their counterparts. Our model equation 5.15 can be viewed as a similar model to equation 6.4 that standard deviation is modeled instead of its square. But since absolute values of aggregate returns are used in our model, the model is not linear, and detailed analytical work becomes more difficult than other models.

It is interesting to note that almost all of the models discussed above and many others in the ARCH family have the same order on both sides of the equations: if σ_t is used, the right hand side would be the linear combinations of first order of returns and standard deviation; if σ_t^2 is used, the right hand side would be squared values of returns and standard deviations. Linear term is only incorporated into σ_t^2 representations when leverage effect is taken into account. This equality in orders appears to come from the apparent mathematical symmetry rather than practical significance.

On the other hand, when our model was derived, since there was a full behavioral picture behind, no such restriction in 'mathematical symmetry' was enforced. Hence the original model equation 5.14 has both quadratic term and linear term. This breaking of order symmetry may be the cause for $\delta = 1.43$ in equation 6.23. More specifically, $1 < \delta < 2$ may be a consequence of the fact that the representation of σ_t^2 contains both linear and quadratic terms.

6.8 Short summary

There are still many other variations of ARCH we have overlooked in the above discussion. It does not mean they are not as meaningful as the ones discussed here, but most of them may share similar motivation and form. Several points can be made about the existing ARCH models as compared to ours in equation 5.15.

1. While fractionally integrated models are able to produce long memory and fat tail with mathematical formulation, their success could be due to the coincidental similarity to the mechanism proposed in our model. In our model we are able to give some clear behavioral interpretation to this type of formulation.

2. According to our model, the reason behind $0 < d^* < 1$ in FIGARCH, and $0 < d^{**} < 0.5$ behind model equation 6.4 are due to the fact that technical strategy is applied at different investment horizons with power-law decay exponent $1 < d < 1.5$.

3. Mathematically, the conditional variance depends on a limited number of past returns, such that the time series can have a finite unconditional variance. The reason behind is that technical traders have finite investment horizons much shorter than fundamental investors. i.e. M has a cutoff corresponding to the limited holding period of technical investors.

4. η_t is time dependent due to opinion convergence, rather than i.i.d., which has been assumed in most ARCH type models. In particular, the conditional distribution of η_t has a larger kurtosis when the past volatility is big.

5. The dependence of conditional volatility on past returns may not be the simple quadratic form suggested by QARCH or HARCH models. Indeed, the authors themselves have argued that the quadratic form is a compromise of mathematical tractability. The form suggested by our model performs equally well, if not better.

These suggest a common mechanism hidden inside some of the ARCH models that produce long memory and fat tail, yet an explicit explanation is given in our model with a similar but different construction.

Chapter 7

Conclusion

Agent-based modeling of economic system is attracting increasing attention in the past decade, and many envision this trend as a necessity due to the increasing complexity of modern economies [106–108]. Physicists are not among the first to adopt such an approach to study economic systems, but their contributions are significant. One of the challenges in the field is to identify the behaviors of individual entities to make realistic models. In the recent decade, the abundance of market data allows researchers to indirectly decipher agent behaviors to some extend, and the growing field of behavioral economics/finance provides invaluable input for model construction.

In financial market, agent-based models can help the understanding of market dynamics at a microscopic level, and explain the macroscopic market fluctuations from the microscopic interactions. While qualitative understanding of market fluctuations has been achieved in the past, producing the stylized facts with quantitative accuracy remained a challenge. This work has made one step closer to this goal. Our agent-based model started from empirical behaviors and was partially calibrated on market data. It is able to capture the tail behavior of return distributions and number of trade distributions. The result has been shown to be robust with a realistic sets of parameter values, indicating a possible dominant mechanism behind market dynamics and statistics. At a day to day level, market fluctuation is dictated by the behavior of technical traders, and the fat tail in returns may correspond to the periods they converge in their trading opinions. From the stochastic volatility model extended from the agent-based model, the volatility clustering effect, or long memory, could be attributed to heterogeneous investment horizons of the agents. This relation can be quantitatively analyzed and empirical proven as demonstrated in this thesis, providing a solid support for the validity of the model.

Another contribution by this work is a possible fine-detail interpretation of general ARCH framework, as well as the introduction of a new stochastic volatility model. Over the years, financial statisticians and market practitioners have developed many variations of the ARCH model, aiming to produce all the major stylized facts including fat tail and long memory. From a mathematical/statistical point of view, the best model should be simple with very few parameters to calibrate, yet can produce return statistics with quantitative accuracy. Some models have come close to this 'holy grail', if it does exists, and the endeavor is still ongoing. As the stochastic volatility model proposed in this thesis originates outright from an agent-based model, and every detail of its construction carries a clear behavioral interpretation, it indirectly explains the meaning its ARCH-like formulation. It is worth noting that the approach here is a bottom-up approach, which starts from specific agent behaviors to a macroscopic model. This is in contrast to many other stochastic volatility models, which are constructed from the macroscopic statistical patterns. It is perhaps not a coincidence that this stochastic volatility model shares some similar traits to certain ARCH family models including HARCH and QARCH, yet we managed to give mathematical analogies to them, and explained why those models work well.

Agent-based and stochastic modeling of financial market have mostly been dealt with on separate paths, and here a link has been constructed between these two approaches, with some success in quantitative accuracy. The agent-based model and the derived stochastic volatility model differ in construction but share the same mechanism of opinion convergence among technical traders. While the agent-based model singles out the dominant market mechanism, the stochastic volatility model allows a detailed analytical study. Both approaches allow their parameter values to be retrieved from market data with clear behavioral interpretations, thus allowing an in-depth study of this highly complex system of financial market.

While the models in this thesis are relatively simple with very few assumptions, the real market place is much more complicated and behaviors of agents are definitely richer than what has been proposed here. Hence there are still grounds for improvements on this work and they can be future research projects. In particular, the leverage effect described in chapter 1 can be examined by introducing buy/sell asymmetry in the model. This can be achieved by examining the statistics of buyer/seller initiated trades on a daily basis. One particular shortcoming of the work is that it ignores the role of high frequency trading (HFT) at a intraday level. How much HFT influence daily price fluctuations is a big question. It is possible that HFT only affects price at intraday level, since the it focuses on short term returns, and any other algorithmic trading with holding period longer than one day can be considered as technical analysis, therefore lies within the consideration of our work.

While the agent-based model is based on the assumption of adaptive agent behaviors, the process of adaptation can be further examined. Market is an evolving system with players adapting their strategies constantly, and adaptive market hypothesis [83] is becoming accepted by researchers. The models assume certain invariant property (the dominance of one strategy at any point in time) in the evolving market place. How does the invariance come about lies within the nature of human beings. It could be greed, fear, herding and others, but the exact process of forming such invariance can be further investigated in greater detail.

Bibliography

[1] B. Mandelbrot. The variation of certain speculative prices. *Journal of business*, pages 394–419, 1963.

[2] P. Gopikrishnan, M. Meyer, L.A.N. Amaral, and H.E. Stanley. Inverse cubic law for the distribution of stock price variations. *The European Physical Journal B-Condensed Matter and Complex Systems*, 3(2):139–140, 1998.

[3] V. Plerou and H.E. Stanley. Tests of scaling and universality of the distributions of trade size and share volume: Evidence from three distinct markets. *Physical Review E*, 76(4):046109, 2007.

[4] R.F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007, 1982.

[5] R.N. Mantegna and H.E. Stanley. Scaling behaviour in the dynamics of an economic index. *Nature*, 376(6535):46–49, 1995.

[6] T. Lux. The stable paretian hypothesis and the frequency of large returns: an examination of major german stocks. *Applied financial economics*, 6(6):463–475, 1996.

[7] V. Plerou, P. Gopikrishnan, L.A.N. Amaral, M. Meyer, and H.E. Stanley. Scaling of the distribution of price fluctuations of individual companies. *Physical Review E*, 60(6):6519, 1999.

[8] P. Gopikrishnan, V. Plerou, L.A.N. Amaral, M. Meyer, and H.E. Stanley. Scaling of the distribution of fluctuations of financial market indices. *Physical Review E*, 60(5):5305, 1999.

[9] V. Plerou and H.E. Stanley. Stock return distributions: Tests of scaling and universality from three distinct stock markets. *Physical Review E*, 77(3):037101, 2008.

[10] T. Kaizoji. A precursor of market crashes: Empirical laws of japan's internet bubble. *The European Physical Journal B-Condensed Matter and Complex Systems*, 50(1):123–127, 2006.

[11] R.K. Pan and S. Sinha. Inverse-cubic law of index fluctuation distribution in indian markets. *Physica A: Statistical Mechanics and its Applications*, 387(8):2055–2065, 2008.

[12] F. Lillo, J.D. Farmer, and R.N. Mantegna. Master curve for price-impact function. *Nature*, 421(129):176–190, 2003.

[13] M. Loretan and P.C.B. Phillips. Testing the covariance stationarity of heavy-tailed time series: An overview of the theory with applications to several financial datasets. *Journal of empirical finance*, 1(2):211–248, 1994.

[14] I.N. Lobato and N.E. Savin. Real and spurious long-memory properties of stock-market data. *Journal of Business & Economic Statistics*, 16(3):261–268, 1998.

[15] É. Rácz, Z. Eisler, J. Kertész, et al. Comment on" tests of scaling and universality of the distributions of trade size and share volume: evidence from three distinct markets". *Physical review. E, Statistical, nonlinear, and soft matter physics*, 79(6 Pt 2):068101–discussion, 2009.

[16] M.M. Dacorogna, U.A. Müller, R.J. Nagler, R.B. Olsen, and O.V. Pictet. A geographical model for the daily and weekly seasonal volatility in the foreign exchange market. *Journal of International Money and Finance*, 12(4):413–438, 1993.

[17] Y. Liu, P. Cizeau, M. Meyer, C.K. Peng, and H. Eugene Stanley. Correlations in economic time series. *Physica A: Statistical Mechanics and its Applications*, 245(3):437–440, 1997.

[18] P. Cizeau, Y. Liu, M. Meyer, C.K. Peng, and H. Eugene Stanley. Volatility distribution in the s&p500 stock index. *Physica A: Statistical Mechanics and its Applications*, 245(3):441–445, 1997.

[19] R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001.

[20] J. Hasbrouck. Measuring the information content of stock trades. *Journal of Finance*, pages 179–207, 1991.

[21] F. Lillo and J.D. Farmer. The long memory of the efficient market. *Studies in Nonlinear Dynamics & Econometrics*, 8(3), 2004.

[22] J.P. Bouchaud, J.D. Farmer, and F. Lillo. How markets slowly digest changes in supply and demand. *Handbook of Financial Markets: Dynamics and Evolution*, 1:57, 2009.

[23] J.P. Bouchaud, Y. Gefen, M. Potters, and M. Wyart. Fluctuations and response in financial markets: the subtle nature of randomprice changes. *Quantitative Finance*, 4(2):176–190, 2004.

[24] X. Gabaix, P. Gopikrishnan, V. Plerou, and H.E. Stanley. A theory of power-law distributions in financial market fluctuations. *Nature*, 423(6937):267–270, 2003.

[25] P. Gopikrishnan, V. Plerou, X. Gabaix, and H.E. Stanley. Statistical properties of share volume traded in financial markets. *Physical Review E*, 62(4):4493–4496, 2000.

[26] M.M. Meerschaert and H.P. Scheffler. A simple robust estimation method for the thickness of heavy tails. *Journal of statistical planning and inference*, 71(1):19–34, 1998.

[27] V. Plerou, P. Gopikrishnan, L.A. Nunes Amaral, X. Gabaix, and H. Eugene Stanley. Economic fluctuations and anomalous diffusion. *Physical Review E*, 62(3):3023–3026, 2000.

[28] Anirban Chakraborti, Ioane Muni Toke, Marco Patriarca, and Frederic Abergel. Econophysics: Empirical facts and agent-based models. 09 2009.

[29] T. Lux and M. Marchesi. Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature*, 397(6719):498–500, 1999.

[30] G. Zumbach. Time reversal invariance in finance. *Quantitative Finance*, 9(5):505–515, 2009.

[31] R. Chicheportiche and J.P. Bouchaud. The fine-structure of volatility feedback. *arXiv preprint arXiv:1206.2153*, 2012.

[32] L.R. GLOSTEN. Is the electronic open limit order book inevitable? *The Journal of Finance*, 49(4):1127–1161, 1994.

[33] J. Hasbrouck. *Empirical market microstructure: The institutions, economics, and econometrics of securities trading.* Oxford University Press, USA, 2007.

[34] D. Challet and R. Stinchcombe. Analyzing and modeling 1+ 1d markets. *Physica A*, 300(1):285–299, 2001.

[35] J.P. Bouchaud, M. Mézard, and M. Potters. Statistical properties of stock order books: empirical results and models. *Quantitative Finance*, 2(4):251–256, 2002.

[36] B. BIAIS, P. HILLION, and C. SPATT. An empirical analysis of the limit order book and the order flow in the paris bourse. *The Journal of Finance*, 50(5):1655–1689, 1995.

[37] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

[38] G. Bottazzi, G. Dosi, and I. Rebesco. Institutional architectures and behavioral ecologies in the dynamics of financial markets. *Journal of Mathematical Economics*, 41(1):197–228, 2005.

[39] P. Pellizzari and A.D. Forno. A comparison of different trading protocols in an agent-based market. *Journal of Economic Interaction and Coordination*, 2(1):27–43, 2007.

[40] A.W. Phillips. Mechanical models in economic dynamics. *Economica*, 17(67):283–305, 1950.

[41] M. Levy, H. Levy, and S. Solomon. A microscopic model of the stock market: cycles, booms, and crashes. *Economics Letters*, 45(1):103–111, 1994.

[42] R. Cont and J.P. Bouchaud. Herd behavior and aggregate fluctuations in financial markets. *Macroeconomic dynamics*, 4(2):170–196, 2000.

[43] R. Cont. Volatility clustering in financial markets: Empirical facts and agent-based models. *Long memory in economics*, page 289, 2007.

[44] S. Mike and J.D. Farmer. An empirical behavioral model of liquidity and volatility. *Journal of Economic Dynamics and Control*, 32(1):200–234, 2008.

[45] T. Lux and M. Marchesi. Volatility clustering in financial markets: A microsimulation of interacting agents. *International Journal of Theoretical and Applied Finance*, 3(04):675–702, 2000.

[46] E. Egenter, T. Lux, and D. Stauffer. Finite-size effects in monte carlo simulations of two stock market models. *Physica A: Statistical Mechanics and its Applications*, 268(1):250–256, 1999.

[47] S. Alfarano and T. Lux. A noise trader model as a generator of apparent financial power laws and long memory. *Macroeconomic Dynamics*, 11:80, 2007.

[48] D. Stauffer. Percolation models of financial market dynamics. *Advances in Complex Systems*, 4(01):19–27, 2001.

[49] LR Da Silva and D. Stauffer. Ising-correlated clusters in the cont-bouchaud stock market model. *Physica A: Statistical Mechanics and its Applications*, 294(1):235–238, 2001.

[50] S. Bornholdt. Expectation bubbles in a spin model of markets: Intermittency from frustration across scales. *International Journal of Modern Physics C*, 12(05):667–674, 2001.

[51] T. Kaizoji, S. Bornholdt, and Y. Fujiwara. Dynamics of price and trading volume in a spin model of stock markets with heterogeneous agents. *Physica A: Statistical Mechanics and its Applications*, 316(1):441–452, 2002.

[52] I. Giardina and J.P. Bouchaud. Bubbles, crashes and intermittency in agent based market models. *The European Physical Journal B-Condensed Matter and Complex Systems*, 31(3):421–437, 2003.

[53] I. Zovko and J.D. Farmer. The power of patience: a behavioural regularity in limit-order placement. *Quantitative finance*, 2(5):387–392, 2002.

[54] J.P. Bouchaud, J. Kockelkoren, and M. Potters. Random walks, liquidity molasses and critical response in financial markets. *Quantitative finance*, 6(02):115–123, 2006.

[55] B. LeBaron, W.B. Arthur, and R. Palmer. Time series properties of an artificial stock market. *Journal of Economic Dynamics and Control*, 23(9):1487–1516, 1999.

[56] D. Challet and Y.C. Zhang. Emergence of cooperation and organization in an evolutionary game. *Physica A: Statistical Mechanics and its Applications*, 246(3):407–418, 1997.

[57] W.A. Brock and C.H. Hommes. Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *Journal of Economic dynamics and Control*, 22(8-9):1235–1274, 1998.

[58] J. Arifovic and R. Gencay. Statistical properties of genetic learning in a model of exchange rate. *Journal of Economic Dynamics and Control*, 24(5):981–1005, 2000.

[59] https://wrds-web.wharton.upenn.edu/wrds/.

[60] http://finance.yahoo.com/.

[61] R.T. Baillie, T. Bollerslev, and H.O. Mikkelsen. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 74(1):3–30, 1996.

[62] B.M. Hill. A simple general approach to inference about the tail of a distribution. *The annals of statistics*, pages 1163–1174, 1975.

[63] B. Efron and R.J. Tibshirani. *An introduction to the bootstrap*, volume 57. Chapman & Hall/CRC, 1994.

[64] L. Stokes. Introduction to variance estimation. *Journal of the American Statistical Association*, 103(483):1324–1325, 2008.

[65] F. Wang, S.J. Shieh, S. Havlin, and H.E. Stanley. Statistical analysis of the overnight and daytime return. *Physical Review E*, 79(5):056109, 2009.

[66] R.A. Wood, T.H. McInish, and J.K. Ord. An investigation of transactions data for nyse stocks. *Journal of Finance*, 40(3):723–39, 1985.

[67] L. Harris. A transaction data study of weekly and intradaily patterns in stock returns. *Journal of financial economics*, 16(1):99–117, 1986.

[68] L. Zhao, G. Yang, W. Wang, Y. Chen, JP Huang, H. Ohashi, and H.E. Stanley. Herd behavior in a complex adaptive system. *Proceedings of the National Academy of Sciences*, 108(37):15058–15063, 2011.

[69] L. Feng, B. Li, B. Podobnik, T. Preis, and H.E. Stanley. Linking agent-based models and stochastic models of financial markets. *Proceedings of the National Academy of Sciences*, 109(22):8388–8393, 2012.

[70] B.R. Brown. *Chasing the Same Signals: How Black-box Trading Influences Stock Markets from Wall Street to Shanghai*. Wiley, 2010.

[71] L. Menkhoff. The use of technical analysis by fund managers: International evidence. *Journal of Banking & Finance*, 34(11):2573–2586, 2010.

[72] M.P. Taylor and H. Allen. The use of technical analysis in the foreign exchange market. *Journal of international Money and Finance*, 11(3):304–314, 1992.

[73] Y.W. Cheung and C.Y.P. Wong. A survey of market practitioners views on exchange rate dynamics. *Journal of International Economics*, 51(2):401–419, 2000.

[74] H. Allen and M.P. Taylor. Charts, noise and fundamentals in the london foreign exchange market. *The Economic Journal*, 100(400):49–59, 1990.

[75] L. Menkhoff. The noise trading approachquestionnaire evidence from foreign exchange. *Journal of International Money and Finance*, 17(3):547–564, 1998.

[76] Y.W. Cheung and M.D. Chinn. Currency traders and exchange rate dynamics: a survey of the us market. *Journal of International Money and Finance*, 20(4):439–471, 2001.

[77] Z. Eisler and J. Kertesz. Liquidity and the multiscaling properties of the volume traded on the stock market. *EPL (Europhysics Letters)*, 77(2):28001, 2007.

[78] HE Hurst. Long-term capacity storage of reservoirs. *Trans Am Soc Civil Eng*, 116:770–799, 1951.

[79] S. Alfarano and T. Lux. A minimal noise trader model with realistic time series properties. Technical report, Economics working paper/Christian-Albrechts-Universität Kiel, Department of Economics, 2006.

[80] S. Alfarano and M. Milaković. Should network structure matter in agent-based finance? 2007.

[81] J. Duffy and N. Feltovich. Does observation of others affect learning in strategic environments? an experimental study. *International Journal of Game Theory*, 28(1):131–152, 1999.

[82] H.P. Boswijk, C.H. Hommes, and S. Manzan. Behavioral heterogeneity in stock prices. *Journal of Economic Dynamics and Control*, 31(6):1938–1970, 2007.

[83] A. Lo. The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. *Journal of Portfolio Management, Forthcoming*, 2004.

[84] V. Plerou, P. Gopikrishnan, X. Gabaix, and H.E. Stanley. Quantifying stock-price response to demand fluctuations. *Physical Review E*, 66(2):027104, 2002.

[85] B. Podobnik, D. Horvatic, A.M. Petersen, and H.E. Stanley. Cross-correlations between volume change and price change. *Proceedings of the National Academy of Sciences*, 106(52):22079–22084, 2009.

[86] S. Alfarano, T. Lux, and F. Wagner. Time variation of higher moments in a financial market with heterogeneous agents: An analytical approach. *Journal of Economic Dynamics and Control*, 32(1):101–136, 2008.

[87] A. Irle, J. Kauschke, T. Lux, and M. Milaković. Switching rates and the asymptotic behavior of herding models. *Advances in Complex Systems*, 14(03):359–376, 2011.

[88] T.G. Andersen and T. Bollerslev. Intraday periodicity and volatility persistence in financial markets. *Journal of empirical finance*, 4(2):115–158, 1997.

[89] R.N. Mantegna and H.E. Stanley. *Introduction to econophysics: correlations and complexity in finance*. Cambridge University Press, 1999.

[90] M.M. Dacorogna. *An introduction to high-frequency finance*. Academic Pr, 2001.

[91] R. Gençay, F. Selçuk, and B. Whitcher. Differentiating intraday seasonalities through wavelet multi-scaling. *Physica A: Statistical Mechanics and its Applications*, 289(3):543–556, 2001.

[92] Y. Liu, P. Gopikrishnan, H.E. Stanley, et al. Statistical properties of the volatility of price fluctuations. *Physical Review E*, 60(2):1390, 1999.

[93] T. Lux. The limiting extremal behaviour of speculative returns: an analysis of intra-daily data from the frankfurt stock exchange. *Applied Financial Economics*, 11(3):299–315, 2001.

[94] Y. Ashkenazy, P.C. Ivanov, S. Havlin, C.K. Peng, A.L. Goldberger, and H.E. Stanley. Magnitude and sign correlations in heartbeat fluctuations. *Physical Review Letters*, 86(9):1900–1903, 2001.

[95] B.G. Malkiel and E.F. Fama. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417, 1970.

[96] F. Black and M. Scholes. The pricing of options and corporate liabilities. *The journal of political economy*, pages 637–654, 1973.

[97] Y. Schwarzkopf and J.D. Farmer. Empirical study of the tails of mutual fund size. *Physical Review E*, 81(6):066113, 2010.

[98] L. Borland and J.P. Bouchaud. On a multi-timescale statistical feedback model for volatility fluctuations. *arXiv preprint physics/0507073*, 2005.

[99] R.F. ENGLE and V.K. NG. Measuring and testing the impact of news on volatility. *The Journal of Finance*, 48(5):1749–1778, 1993.

[100] J.M. Zakoian. Threshold heteroskedastic models. *Journal of Economic Dynamics and control*, 18(5):931–955, 1994.

[101] C.W.J. Granger and Z. Ding. Some properties of absolute return: An alternative measure of risk. *Annales d'Economie et de Statistique*, pages 67–91, 1995.

[102] U.A. Müller, M.M. Dacorogna, R.D. Davé, R.B. Olsen, O.V. Pictet, and J.E. von Weizsäcker. Volatilities of different time resolutionsanalyzing the dynamics of market components. *Journal of Empirical Finance*, 4(2):213–239, 1997.

[103] P.E. Lynch and G.O. Zumbach. Market heterogeneities and the causal structure of volatility. *Quantitative Finance*, 3(4):320–331, 2003.

[104] E. Sentana. Quadratic arch models. *The Review of Economic Studies*, 62(4):639–661, 1995.

[105] Z. Ding, C.W.J. Granger, and R.F. Engle. A long memory property of stock market returns and a new model. *Journal of empirical finance*, 1(1):83–106, 1993.

[106] J.P. Bouchaud. Economics needs a scientific revolution. *Nature*, 455(7217):1181–1181, 2008.

[107] T. Lux and F. Westerhoff. Economics crisis. *Nature Physics*, 5(1):2–3, 2009.

[108] J.D. Farmer and D. Foley. The economy needs agent-based modelling. *Nature*, 460(7256):685–686, 2009.